Inductive and Deductive Reasoning in Obsessive-Compulsive Disorder

Janice Hui-Lin Liew

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy/ Master of Psychology (Clinical)

School of Psychology

Faculty of Science

August 2015

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed:

Date:

Acknowledgements

First and foremost, I would like to thank my supervisor, **Professor Brett Hayes**, for his invaluable guidance and support at every stage of this research program. His feedback in the development and implementation of this research has been indispensable, and I am grateful for his steady, optimistic approach to supervision. I also greatly appreciate the generosity he has always shown with his time and knowledge. My deep gratitude also goes to my co-supervisor, **Associate Professor Jessica Grisham**, for her consistent encouragement, advice, and expertise in all things related to obsessive-compulsive disorder. Thank you also to **Dr Naomi Sweller**, whose statistical advice was extremely helpful.

I would also like to thank past and present members of the Cognition lab for their recommendations and advice. I would like to especially thank **Melissa Lim, Kelly Jones,** and **Kate Blundell** for their assistance with programming and experimentation, in addition to the good cheer that they always brought to the lab.

I am indebted to all my colleagues on level 14 for their unceasing and unconditional support on a moment to moment basis. In particular, I would like to thank **Michelle** for the innumerable ways in which she has patiently encouraged my development as a clinician, researcher, and more importantly, as a person. My gratitude also goes to **Donna,** my running mate since the dawn of this UNSW journey, who has always given me new perspectives and has been all the company I have ever needed. My sincere thanks also to **Dom, Jade, Hui, Jess,** and **Miri,** for all the questions answered, for all the times you've spurred me on, and for all the joy you've brought to this process. I honestly do not know if I could have done any of this without all of you.

My deepest gratitude goes to my friends and family, who have always been patient with my busyness and unwavering in their care. Particular thanks goes to **Cathy,** for her inspirational cheerleading, meme-finding, and baking; to **Jess,** for her constant affirmations and for her unfailing ability to make me laugh; to **Tab,** my best reminder that life exists outside of the PhD; to **Mel,** whose friendship is a firm foundation; and to my **small group,** for their persistence in prayer. Thanks also to **Janine, Sienna, Tina, Dani, Shirley,** and **Fei,** who all have (at various times) pre-emptively changed my title, and whose encouragement has never been bound by proximity. Additionally, to my oldest friends, **EBCT**, I look forward to our dog-filled house!

And to **Mum, Dad,** and **Nat,** thank you for the boundless ways in which you have supported and provided for me, and for always keeping me grounded. In large part, I owe the completion of this thesis to you.

Most of all to my God, **Jesus Christ**, to whom I owe a debt that I can never repay, thank you for Your endless mercies and Your redeeming grace.

And finally, to those who have persistently asked after my progress all these years and who have encouraged me in so many different ways (there are too many of you to name!) – with my whole heart, I thank you.

Abstract

This body of research investigated how individuals with high levels of obsessive-compulsive disorder (OCD) symptomatology or diagnosed OCD reason when presented with arguments that lead to certain conclusions (deductive reasoning) or probabilistic conclusions (inductive reasoning). Previous research has suggested a "spared deduction, impaired induction" account of OCD, where OCD patients are impaired on tasks of inductive reasoning but are unimpaired on most tasks of deductive reasoning. This thesis sought to test this account by utilising tasks adapted from the contemporary reasoning literature that allowed for a direct comparison between deductive and inductive reasoning. In numerous experiments, participants were presented with a common argument set and asked to make judgments about the deductive validity or inductive plausibility of the arguments. Deductive deficits in sensitivity to argument validity were noted in analogue samples high in OCD symptomatology and in individuals with diagnosed OCD on class-inclusion problems (Experiments 1, 2 and 8), but not when validity was manipulated via other logical forms (i.e., modus ponens, conjunctive syllogisms, or categorical syllogisms; Experiments 6 and 7). Further, individuals high in OCD symptoms were unimpaired in their use of background causal knowledge (Experiment 6) and sample size (premise monotonicity; Experiments 1 and 8) when making inductive judgments. However, both analogue samples high in OCD symptoms and an OCD patient group exhibited a reduced sensitivity to the implications of premise diversity in induction (Experiments 2, 3 and 8). This may be partially explained by differences in the way that individuals high and low in OCD symptoms perceive similarity between the premises (Experiment 5). Importantly, these deductive and inductive deficits were found across both OCD-neutral and OCD-relevant item content and were not due to group differences in general ability. These results challenge the “spared deduction, impaired induction” account of OCD. Cognitive therapies for OCD may be improved by incorporation of direct instruction about the implications of diverse evidence and how to distinguish between simple valid and invalid arguments.

Table of Contents

List of Tables

List of Figures

CHAPTER 1 - General Introduction

Overview

This thesis examined reasoning processes in obsessive-compulsive disorder (OCD). Specifically, the thesis investigated how individuals high in OCD symptomatology (Chapters 2 to 4) or diagnosed OCD (Chapter 5) reason when presented with arguments that lead to conclusions that could be made with certainty (deductive reasoning) or that were probabilistic (inductive reasoning). Previous research has suggested a “spared deduction, impaired induction” account of OCD, in which differences in performance have been noted between OCD patients and controls on a variety of inductive reasoning tasks, but not on most deductive reasoning tasks. Primarily, this thesis aimed to clarify whether this account reflected the true state of inductive and deductive reasoning processes in OCD. Notably, the thesis extended existing research by utilising tasks adapted from the contemporary reasoning literature that allowed for a direct comparison between inductive and deductive processes.

The experiments in this thesis also provided a more comprehensive review of OCD-related deficits in inductive reasoning by examining two different types of induction: induction based on background knowledge, and induction based on domain-general sampling heuristics such as premise monotonicity and premise diversity. Importantly, in the majority of experiments, reasoning in high and low OCD symptom populations was examined whilst controlling for the effects of general ability. Additionally, this thesis examined whether impairments in reasoning were specific to OCD-relevant content or global across both OCD-neutral and OCD-relevant content.

To better understand the clinical and theoretical context for the current experiments, this chapter reviews the phenomenology of OCD, the role of cognitive deficits in the aetiology and maintenance of OCD symptoms, and current approaches to examining deductive and inductive reasoning. Existing research on deductive and inductive reasoning in OCD is then examined. Following this, conceptual and methodological limitations with the extant research are discussed, leading to an examination of how the current thesis aimed to resolve these difficulties to extend current understanding of reasoning in OCD.

Obsessive-Compulsive Disorder

Obsessive-compulsive disorder (OCD) is a mental disorder that affects approximately 3% of the Australian population in any given 12-month period (McEvoy, Grove, & Slade, 2011). The typical onset of symptoms is gradual (American Psychiatric Association, 2013), with the first symptoms often occurring in adolescence. The median age of onset is approximately 19 years (Kessler et al., 2005; McEvoy et al., 2011; Ruscio, Stein, Chiu, & Kessler, 2010). OCD is debilitating, causing significant impairment in everyday function (Koran, Thienemann, & Davenport, 1996; Leon, Portera, & Weissman, 1995), and if left untreated, tends to be chronic in nature, with symptoms fluctuating in severity and often increasing in times of stress (Rasmussen & Eisen, 1992). OCD frequently occurs together with other psychological disorders (e.g., anxiety disorders, mood disorders, eating disorders, and substance abuse disorders), with up to 90% of patients meeting diagnosis for another psychological disorder at some point in their lifetime (Ruscio et al., 2010).

Phenomenology

OCD is characterised by the experience of unwanted and repetitive intrusive thoughts and images (i.e., obsessions) which often compel an individual to perform some distress-reducing behaviour or mental act (i.e., a compulsion). OCD patients generally have some insight into the senselessness or excessiveness of their symptoms, although this level of insight varies greatly from individual to individual (Muller & Roberts, 2005).

Obsessions. Obsessions are defined as recurrent and persistent thoughts, urges or images that are intrusive, unwanted, and usually cause anxiety or distress (Diagnostic and Statistical Manual of Mental Disorders [DSM-5], American Psychiatric Association, 2013). The majority of patients present with more than one obsession (Farrell, Barrett, & Piacentini, 2006; Geller et al., 2001; Rasmussen & Tsuang, 1986), which typically centre on a number of main themes: concerns about preventing possible harm, contamination, the need for symmetry, and unacceptable religious, aggressive, or sexual content (Abramowitz, Franklin, Schwartz, & Furr, 2003). For example, an individual may experience repetitive and unwanted intrusions about the possibility of the house burning down due to a stove being left on or the possibility of contracting disease via physical contact with germs on hands.

Although these intrusions are a key characteristic of OCD, their presence alone is often insufficient to distinguish between OCD patients and the non-clinical population. Indeed, the content and form of normal obsessions (experienced by 80-90% of the general population) can be difficult to distinguish from abnormal obsessions, even by those with relevant clinical experience (Purdon & Clark, 1993; Rachman & de Silva, 1978; Salkovskis & Harrison, 1984). Nevertheless, a number of studies have shown that reliable discrimination between normal and abnormal obsessions can be achieved (Rassin, Cougle, & Muris, 2007; Rassin & Muris, 2007). Clinical intrusions tend to be more frequently re-experienced, contain more violent and bizarre content, and are experienced more often in the absence of explicit triggers (see Berry & Laskey, 2012 for a review). Furthermore, OCD patients are more likely to appraise themselves as being both responsible for, and able to control, their intrusions, which heightens distress and increases the frequency with which the intrusions are experienced (Berry & Laskey, 2012).

Compulsions. Obsessions are often accompanied by compulsions (repetitive behaviours or mental acts) that represent attempts to reduce or neutralise the marked anxiety and distress that the obsessions cause, or alternatively, to prevent some feared negative event occurring (American Psychiatric Association, 2013). For example, in response to the obsession that one might contract a disease due to germs present on the hands, an individual might spend multiple hours a day washing their hands in a highly ritualised manner (e.g., washing different parts of the hands in a particular order and for a particular length of time in order to be considered ‘clean’).

These repetitive behaviours (e.g., hand-washing, ordering, checking) or mental acts (e.g., counting, praying, repeating phrases internally) are typically excessive, time-consuming, and are often not realistically connected to the feared event they were designed to prevent (American Psychiatric Association, 2013). Almost 80% of OCD patients report having both mental and physical compulsions, and the majority of the remaining 20% report experiencing physical compulsions alone (Foa & Kozak, 1995).

Although the diagnostic profile described above accurately describes many OCD patients, it is also widely acknowledged that OCD is a highly individualised disorder. Whilst obsessions and compulsions can be clustered according to the aforementioned thematic concerns (e.g., contamination, doubting, ordering/symmetry), individuals vary considerably in the number and nature of the themes they find concerning (Moritz, Kloss, & Jelinek, 2010).

Cognitive Correlates and Deficits in OCD

Many fundamental cognitive processes, such as attention, executive function, and memory, are postulated to be impaired and/or biased in OCD. Additionally, impaired cognition is also thought to play a key role in maintaining the disorder via the appraisals of one's symptoms. For example, Salkovskis' (1985, 1989, 1999) model of OCD suggests that it is not simply the experience of intrusions, but rather the interpretation that one is responsible for the outcome and/or content of the intrusions, that causes anxiety and corrective neutralising behaviours. Similarly, Rachman (1997, 1998) argues that obsessions are maintained by misinterpretation of the significance of one's intrusions (i.e., interpreting intrusions as personally significant, threatening, or catastrophic). For example, if a person strongly values moral purity, then inappropriate sexual intrusions may be interpreted to mean that the person is morally bankrupt. This is a view of oneself that is not easy to ignore or dismiss, and thus, such intrusions are likely to develop into an obsession.

This section briefly explores a selection of key findings regarding how cognitive deficits in attention, memory, and executive function may play a role in the aetiology and maintenance of OCD. Although it is outside the scope of this thesis to provide a thorough review of these areas, a general understanding of cognition in OCD will assist in an understanding of the role played by reasoning, the higher-order cognitive process that is the focus of this thesis.

Attention and Inhibitory Control

Biases in attention may trigger and maintain intrusions via an enhanced capacity to encode threat-relevant information (e.g., Foa & McNally, 1986; Tata, Leibowitz, Pmnty, Cameron, & Pickering, 1996). For example, on a visual dot-probe task, OCD patients with contamination concerns respond faster to contamination words, whereas anxiety controls respond faster to words related to social anxiety (Tata et al., 1996).

The repetitive thoughts and behaviours that characterise OCD may also reflect an impairment of inhibitory control. For example, obsessions may be conceptualised as a failure to inhibit or shift attention away from distressing intrusions, whilst compulsions may be related to an inability to behaviourally inhibit prepotent responses (Bannon, Gonsalvez, Croft, & Boyce, 2002; Chamberlain, Blackwell, Fineberg, Robbins, & Sahakian, 2005; Penadés et al., 2007). This notion is supported by the difficulty OCD patients experience when attempting to ignore the semantic content of OCD-relevant words on a modified Stroop task (Foa, Illai, McCarthy, Shoyer, & Murdock, 1993; Lavy, van Oppen, & van den Hout, 1994; McNally et al., 1994), and their increased errors of commission (i.e., poorer behavioural inhibition) on the go/no-go task (Bannon et al., 2002; Penadés et al., 2007), relative to both anxious and non-clinical controls.

Memory in Relation to Doubt and an Intolerance of Uncertainty

Pathological doubt and an intolerance of uncertainty characterises all forms of OCD (Obsessive Compulsive Cognitions Working Group, 1997; Tolin, Abramowitz, Brigidi, & Foa, 2003). OCD patients, for example, will often report serious doubts about whether their hands have been sufficiently washed or whether they have committed particular sins. The pathological doubt that follows such intrusions often prompts compulsive behaviours aimed at reducing doubt and increasing certainty (Carr, 1974; Guidano & Liotti, 1983; Kozak, Foa, & McCarthy, 1987).

Neuropsychiatric models of OCD propose that the repetitive nature of these compulsions may be partially explained by memory impairment (Deckersbach, Otto, Savage, Baer, & Jenike, 2000). If one cannot remember switching off the stove, then it seems natural to doubt whether the stove was turned off safely. Experimental evidence for this account, however, is not strong (see Muller & Roberts, 2005 for a review). A number of experimental studies have shown that OCD patients often do not differ from controls in recall and recognition accuracy; they do, however, often express less *confidence* in their memories (Hermans, Martens, De Cort, Pieters, & Eelen, 2003; MacDonald, Antony, MacLeod, & Richter, 1997; McNally & Kohlbeck, 1993), and greater dissatisfaction with their level of memory vividness (Constans, Foa, Franklin, & Mathews, 1995).

It is not difficult to see how a reduced memory confidence and a desire to increase memory vividness might compel someone to repeated checking actions. Such checking, however, has a counterproductive effect on memory. In OCD patients, repeated exposure to threat-related objects actually leads to a further reduction in memory confidence not observed in anxious and non-clinical controls (Tolin et al., 2001), and the act of repeated checking decreases trust in both memory accuracy and memory vividness (Boschen & Vuksanovic, 2007; Coles, Radomsky, & Horng, 2006; Radomsky, Gilchrist, & Dassault, 2006; van den Hout & Kindt, 2003a, 2003b).

One further difficulty with the memory-deficit account is that persistent doubting is seen in situations where disconfirming evidence is readily available. That is, the urge to perform compulsions persists even when the OCD patient does not have to rely on their memory for evidence of safety. For example, a checker may become trapped in repetitive cycles of lock-checking, or a washer in cycles of hand-washing, despite the presence of clear perceptual evidence indicating that the door is locked or that one's hands are clean. Instead, the individual continues to engage in compulsive behaviours that seemingly reflect the need to collect more uncertainty-reducing evidence (Rotge et al., 2013; Stern et al., 2013).

Descriptively, such deficits fit with the notion that OCD patients have difficulty reasoning inductively (i.e., drawing uncertain or probabilistic conclusions from available evidence). This parallels the suggestion that impairment in evidential rules for inferring danger may prolong a patient's doubt about the just-checked gas stove (Foa & Kozak, 1985). If OCD patients underweight the contribution of given information, or have a higher threshold for the amount of information necessary to make a decision, then this is likely to contribute to pathological doubts about issues such as safety or contamination. Indeed, OCD patients are more cautious when making decisions, frequently requesting more information or for information to be repeated before making a final choice (e.g., Fear & Healy, 1997; Milner, Beech, & Walker, 1971).

Executive Function

An important component of executive functioning is the ability to flexibly switch attention to different features of task stimuli in response to changing demands. At least superficially, the perseverative and ritualistic behaviour often observed in OCD could be conceptualised as a breakdown in executive flexibility (Olley, Malhi, & Sachdev, 2007). Perseverative behaviour is often examined using the Wisconsin Card Sort Test (Berg, 1948; Grant & Berg, 1948). On this task, participants must determine the correct rule by which to sort cards (e.g., according to colour or number), utilising the feedback (“correct” or “incorrect”) given by the experimenter. Once the participant has correctly sorted 10 consecutive trials, the rule is changed without notifying the participant. From this, a measure of perseveration can be derived: the number of times the participant continues to sort according to the old rule. On this task, OCD patients tend to commit a larger number of perseverative responses relative to matched non-clinical controls (e.g., Lawrence et al., 2006; Lucey et al., 1997; Okasha et al., 2000; Roh, Shin, Kim, Lee, & Kwon, 2005).

The Role of Reasoning in Obsessive-Compulsive Disorder

As suggested above, reasoning is thought to play a role in the aetiology and maintenance of OCD. Reasoning involves drawing inferences (or conclusions) on the basis of known information or evidence (Holyoak & Morrison, 2005). People reason in order to make decisions that they can justify and to create action plans that will facilitate achievement of their goals (Evans, Over, & Manktelow, 1993; Johnson-Laird & Shafir, 1993).

This thesis is primarily concerned with examining two kinds of reasoning in individuals who vary in their levels of OCD symptomatology: 1) reasoning where conclusions necessarily follow (or do not follow) from the given evidence, and 2) reasoning where conclusions are probabilistic (i.e., are more or less likely) given the evidence. In the wider reasoning literature, these approaches correspond to deductive and inductive reasoning respectively (Heit, 2007).

This section briefly reviews contemporary research on deductive and inductive reasoning. To establish a platform for understanding how these reasoning processes might be best studied in OCD, this section explores models of and factors which affect deductive and inductive reasoning. In addition, empirical and methodological issues that arise in the study of these processes are examined.

Deductive Reasoning

Deductive reasoning is the process by which conclusions are generated which either do or do not necessarily follow from given premises. *Deductively valid* arguments lead to conclusions that must be true, as long as the premises are true (Jeffrey, 1981). Consider the following deductive argument that relies on the logical principle of class-inclusion:

All birds lay eggs

All emus are birds

All emus lay eggs

In this example, the information above the line (hereafter referred to as the argument *premises*) is assumed to be true, and the conclusion (below the line in italics) is assessed for deductive validity (i.e., whether the conclusion necessarily follows from the given premises).

Arguments can also be *deductively invalid*, such that conclusions cannot follow solely from the given information. An example of a deductively invalid argument is:

All emus are birds

All emus lay eggs

All birds lay eggs

Note that the conclusion itself may be empirically accurate (all birds do lay eggs), but if not necessitated by the premises (knowing that all emus are birds that lay eggs does not necessarily mean that all birds lay eggs), the argument is deductively invalid. This illustrates an important feature of deductive arguments: validity depends on the logical forms linking the premises to the conclusion, rather than the empirical truth of the argument content.

Deductive reasoning is a vital part of both everyday and scientific reasoning. Without deduction, there would be no formal system of mathematics and no scientific process. Failures to draw deductively valid conclusions can also contribute to mistakes that may result in negative consequences (e.g., incorrectly continuing to use dangerous machinery when warning indicators are signaling to stop; Johnson-Laird, 1999).

Factors that influence deductive judgments. There are a plethora of factors that are postulated to influence deductive judgments on experimental tasks (see Evans, 1982; Johnson-Laird, 1999; Rips, 1994 for reviews). This section focuses on the two factors particularly relevant to the current thesis: argument type and argument content.

Argument type. Deductive arguments can take many forms. Conditional reasoning, for example, focuses on four main inference types laid out in Table 1.1. The form which an argument takes can influence deductive performance. For example, of the valid arguments, modus ponens arguments are easier to reason with, and are therefore endorsed more often than modus tollens arguments (Johnson-Laird, 1999). In addition, although denying the antecedent and affirming the consequent are logical fallacies, they are frequently endorsed, especially when the conclusion is consistent with background knowledge (Oaksford & Chater, 2001).

Table 1.1

Examples of Conditional Reasoning Argument Forms

Argument validity	Argument type	Argument form
Valid	Modus ponens	If A then B A Therefore B
	Modus tollens	If A then B Not B Therefore not A
Invalid	Affirming the consequent	If A then B B Therefore A
	Denying the antecedent	If A then B Not A Therefore not B

Argument content. Formally, deductive reasoning should not be influenced by background beliefs about the empirical truth of argument premises or conclusions, but should depend solely on the application of logical rules. However, the believability of an argument's conclusion does often influence deductive inferences. To illustrate, consider this example of an invalid syllogism (where the conclusion does not necessarily follow from the premises) with a believable conclusion used by Evans, Barston, and Pollard (1983):

No addictive things are inexpensive

Some cigarettes are inexpensive

____

Therefore, some addictive things are not cigarettes

A valid syllogism with an unbelievable conclusion can be created by simply re-ordering the conclusion:

No addictive things are inexpensive

Some cigarettes are inexpensive

____

Therefore, some cigarettes are not addictive

Previous work has shown that just as people more frequently endorse valid than invalid arguments, so too do they more frequently endorse believable (as in the first example) than unbelievable arguments, regardless of logical validity (Evans, 1989; Evans et al., 1983; Newstead & Evans, 1993). This is because people generally have a tendency to confirm, rather than to disconfirm, their prior beliefs (Evans, Over, et al., 1993). This effect of believability is larger on invalid arguments than on valid ones (Newstead, Pollard, Evans, & Allen, 1992).

Theories of deductive performance. Classically, there have been two main ways of conceptualising deductive reasoning: as a system of formal inferential rules (Braine & O'Brien, 1991; Rips, 1994) or the result of inference based on mental models (Johnson-Laird & Byrne, 1991; Polk & Newell, 1995).

Deduction based on formal inferential rules**.** Braine and O'Brien (1991) and Rips (1994) proposed that when reasoning, people convert arguments into premises in logical form, and then follow rules to derive the appropriate conclusions. For example, consider the following premises:

If there is a fox or cat in the toy box, then there is not a pear.

It is false that there is not a fox in the toy box.

Reasoning can now proceed by the application of particular inferential rules (e.g., double negatives are cancelled as a first step, which adds the information that there is a fox in the toy box), leading to the conclusion that there is no pear in the toy box (Braine & O'Brien, 1998).

Deduction based on mental models. Mental models theory postulates that reasoning is based on the formation and manipulation of mental models that are created as representations of situations (Johnson-Laird & Byrne, 1991; Polk & Newell, 1995). Each mental model represents a possibility. For example, consider this argument: there is a circle or there is not a triangle. This argument must be represented by three separate mental models: there is a circle, there is no triangle, and there is a circle but no triangle. Each of these mental models represents a true possibility that captures commonalities in any situation where the circle exists but the triangle does not.

According to mental models theory, if the premises are true then conclusions are possible if at least one mental model can be created from the premises (Bell & Johnson-Laird, 1998). The more models a conclusion includes, the more probable the conclusion becomes (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), and if present in all the possible mental models, then the conclusion is necessary (Johnson-Laird & Byrne, 1991).

There is no consensus regarding which of these models better accounts for deductive reasoning; deduction may rely on one of these processes or be some combination of both (Falmagne & Gonsalves, 1995). However, this is not a theoretical distinction that is central to the current thesis. Of crucial importance is the notion that these aforementioned models of deduction implicitly assume that the processes that underlie deduction are distinct from those that are assumed to underlie inductive reasoning[1].

Inductive Reasoning

Inductive reasoning involves generalising existing information or evidence to novel instances or situations (Hayes, Heit, & Swendsen, 2010). Whilst inductive conclusions may be possible given the evidence (and perhaps even likely), they are never certain or logically necessary. This type of reasoning would be employed if one were trying to predict whether a novel bird can fly. Because flight is a feature of most known bird species, it is likely (but still uncertain) that the feature will generalise to a new member of the species. The strength of this prediction typically comes in *degrees*; this degree of belief can be modelled as some probability, just as induction can be modelled as some rule for evaluating changing probabilities in light of given evidence (Hájek & Hall, 2002).

Because inductive reasoning deals with inferences that are probabilistic or uncertain, it strongly relates to the reasoning employed in everyday life. People frequently make probabilistic predictions based on existing knowledge, such as whether unfamiliar birds can fly, whether roads will be congested when leaving work, or whether the weather will allow for a weekend beach trip. These inferences often inform behaviours, with umbrellas either taken or left at home depending on inductive inferences about the weather.

Inductive reasoning is also related to numerous cognitive activities, including (but not limited to) categorisation, judgments of probability, and decision-making (Heit,

[1] This assumption is contested by Oaksford and Chater (2001, 2007), who argue that a probabilistic Bayesian approach can account for how people reason deductively, just as a probabilistic approach can account for inductive reasoning. For example, conditional reasoning can be interpreted via conditional probability, where the statement ‘birds fly’ suggests a high conditional probability of something flying, given that it is a bird. Such approaches are discussed further in the section comparing single- and dual-process theories of reasoning.

2007). Expansion of scientific knowledge also demands inductive inferences; if deduction alone were used, nothing could be learnt that was not already implicit in the premises (Evans & Over, 2013). At the same time, however, because inductive inferences are not logically valid, at best these inferences can be held with a degree of confidence or subjective probability, but never with certainty.

Experimental studies often use category-based induction problems, which involve making an inference about the properties of a conclusion category based on the known properties of a premise category or categories (Heit, Rotello, & Hayes, 2012). For example, as noted above, a particular property (flight) that applies to eagles and pigeons (and numerous other bird species) may be generalised to an unfamiliar bird, or in the absence of knowledge to the contrary (i.e., knowledge that not all birds fly), may even be generalised to 'all birds'. This argument can be structured in the same way as deductive syllogisms:

All eagles can fly

All pigeons can fly

This unfamiliar bird can fly

Factors that influence inductive judgments. A review of research on induction indicates there are at least two distinct factors that influence category-based inductive inferences (see Bright & Feeney, 2014 for an alternate approach to distinguishing between factors that influence induction; see Hayes et al., 2010 for a detailed review). The first factor is the use of general heuristics for assessing probabilistic evidence (Heit, 2000; Osherson, Smith, Wilkie, López, & Shafir, 1990). These heuristics are based on the characteristics of any given sample (e.g., the size or diversity of the sample), and represent general rules of thumb for deciding whether known information is readily generalisable. It is thought that such general rules are used as default principles to guide induction when one has relatively little specific knowledge about the categories involved and the property being generalised (Hayes et al., 2010; Shafto & Coley, 2003).

The second factor that influences inductive judgments is existing background knowledge, which often overrides the use of general heuristics (Medin, Coley, Storms, & Hayes, 2003; Rehder, 2006; Shafto & Coley, 2003). In this context, ‘background knowledge’ is taken to include domain expertise, knowledge of causal relations between premises and the conclusion, and specific knowledge about the to-be-generalised properties.

Inductive sampling heuristics. Numerous inductive heuristics have been identified, falling under two main categories: heuristics to generalise information from a single case (e.g., premise similarity and typicality), and heuristics to generalise from multiple cases (e.g., premise monotonicity and diversity). People are most likely to use these heuristics (information about the characteristics of the sample) when generalising unfamiliar or “blank” properties (e.g., ‘has Property X’, ‘has ileal veins’). A number of the main inductive sampling heuristics are summarised below.

Premise-conclusion similarity. The more similar a premise is to the conclusion category, the stronger the inductive argument. For example, participants are more likely to generalise the property ‘uses serotonin as a neurotransmitter’ from robins to sparrows than from robins to geese, due to higher perceived similarity between robins and sparrows (Heit, 2000; Osherson et al., 1990).

Premise typicality. The more representative or typical a premise is of a more general conclusion category, the stronger the inductive argument. For example, knowing that sparrows (a typical bird) have a ‘higher potassium concentration in their blood than humans' is stronger evidence that all birds share this property, compared to knowing that penguins have this property (Heit & Rotello, 2005; Osherson et al., 1990).

Premise monotonicity (sample size). Inductive arguments based on many premises that share a property (e.g., donkeys, wolves, goats, monkeys, and dogs having 'Property X') are seen as a stronger basis for property generalisation than arguments with fewer positive premises (e.g., donkeys and wolves having 'Property X'; Osherson et al., 1990). That is, the strength or plausibility of an inductive conclusion tends to increase with the number of instances of positive evidence observed. Such premise monotonicity effects have been repeatedly demonstrated in studies of inductive reasoning with non-clinical participants (e.g., Choi, Nisbett, & Smith, 1997; Feeney, 2007; Heit & Rotello, 2012; McDonald, Samuels, & Rispoli, 1996; Osherson et al., 1990; Rotello & Heit, 2009).

Premise diversity. When making inductive judgments, an individual must consider how well particular pieces of evidence support their judgments. One factor that can increase the strength of a conclusion is when the supporting evidence is diverse in nature. All things being equal, more diverse evidence is usually seen as a stronger basis for inductive generalisation than less diverse evidence. This makes intuitive sense: when testing a hypothesis, one should search for additional supporting evidence that is dissimilar to the evidence already on hand. However, this effect is limited to cases where all premises belong to the same superordinate category.

Experimentally, Osherson et al. (1990) demonstrated this principle by asking participants to choose whether a diverse argument (e.g., 'cows have Property X' and 'mice have Property X') or non-diverse argument (e.g., 'cows have Property X' and 'horses have Property X') was stronger evidence for a particular conclusion (e.g., 'mammals have Property X'). Reliably, participants judged the diverse arguments to be stronger, demonstrating sensitivity to the greater diversity between cows and mice compared to cows and horses.

Diverse evidence is also more frequently preferred over non-diverse evidence in hypothesis testing (López, 1995). For example, participants were given a fact about a mammal (e.g., ‘lions have a dolichoid cerebellum’) and were asked which additional mammal (e.g., leopards or goats) they would test to see whether all mammals possessed this property. Goats were more frequently chosen as the to-be-tested category, demonstrating that participants viewed dissimilar evidence as a stronger test of the conclusion.

Diversity effects have been repeatedly demonstrated (see Heit, 2000 for a review; and Kim & Keil, 2003; Kincannon & Spellman, 2003 for more recent evidence), and occur because diverse premises bring general and highly projectable features to mind (Feeney & Heit, 2011). In contrast, non-diverse premises prime idiosyncratic and less projectable features.

The influence of background knowledge on inductive reasoning. An individual’s background knowledge influences the inductive generalisations they are willing to make. Background knowledge can take a variety of forms, including domain expertise, property knowledge, and knowledge of causal relations.

Domain expertise and taxonomic knowledge. People who have particular domain expertise use this knowledge when generalising properties within the realm of their expertise (López, Atran, Coley, Medin, & Smith, 1997; Proffitt, Coley, & Medin, 2000). For example, López et al. (1997) compared the use of premise diversity in US undergraduates and the indigenous Itza Maya from Central America. When reasoning about arguments that involved biological kinds, the US undergraduates showed the usual sensitivity to argument diversity. In contrast, the Itza Maya favoured conclusions based on non-diverse (i.e., highly similar) premises. The Itza Maya appeared to be using their more extensive knowledge of local biology to draw inferences, superseding the heuristic effect of diversity (e.g., they reasoned that diseases could only be passed to two dissimilar mammals through an infected animal biting both, whereas the non-diverse pair was similar enough for the disease to spread naturally).

Note that experts only use this extra knowledge when properties fall within the realm of their expertise and when they have sufficient time. For example, fishermen only generalised according to ecological reasons for novel marine diseases, and reverted to taxonomic similarity to generalise novel blank properties, whereas novices relied on taxonomic similarity for both kinds of property generalisations (Shafto & Coley, 2003). Additionally, in the face of time pressure, inductions default to being taxonomically based (Shafto, Coley, & Baldwin, 2007). Thus, rather than abandoning the use of similarity altogether, experts use both taxonomic relations and additional knowledge to make their inductive inferences (Hayes et al., 2010; Shafto & Coley, 2003). A relationship between knowledge type and reasoning context has also been observed in non-expert populations, where associative knowledge (e.g., co-occurrence or similarity relationships) is favoured over background knowledge (e.g., taxonomic and causal relationships) when participants are under time pressure or a cognitive load; otherwise, structured knowledge is preferred (Bright & Feeney, 2014).

Property knowledge. Inductive generalisations are also often influenced by knowledge of the to-be-generalised property. Heit and Rubinstein (1994) noted differences between inferences made about anatomical properties and behavioural properties. For example, participants were more likely to generalise anatomical properties between biologically similar animals (e.g., whales and bears both having ‘livers with two chambers that act as one’), whereas they were more likely to generalise behavioural properties between behaviourally matched species (e.g., whales and tuna both ‘travel in a back-and-forth, or zig-zag trajectory’).

Causal knowledge. Knowledge of causal relationships between premise and conclusion categories is one of many types of background knowledge that can affect property induction and categorisation (Keil, 2006; Medin et al., 2003). For example, ‘zebras have Enzyme X, therefore lions have Enzyme X’ may seem inductively strong to an individual who retrieves the prior knowledge that zebras are often dinner options for lions, activating the causal inference that Enzyme X may be transmitted through digestion. Similarly, Rips (2001) demonstrated that inductive conclusions are more likely to be accepted when consistent (rather than inconsistent) with pre-existing causal knowledge, even when neither type of conclusion necessarily followed from the given premises.

Causal knowledge is often preferred over the use of inductive sampling heuristics when making inductive generalisations (Medin et al., 2003; Rehder, 2006, 2009; Rehder & Burnett, 2005). For example, dissimilar premise-conclusion arguments that have salient causal links are perceived to be as strong as, or stronger than, similar premise-conclusion pairs with no apparent causal links (Medin et al., 2003; Rehder, 2006). In addition, Shafto, Kemp, Bonaqitz, Coley, and Tenenbaum (2008) demonstrated that knowledge of causal relations between premises and conclusions can also generate generalisation asymmetries, such that property projection from a cause to an effect (e.g., a property of acorns generalising to squirrels) is seen as stronger than one from an effect to a cause (e.g., a property of squirrels generalising to acorns).

Theories of inductive reasoning.

Similarity-based approaches. There are at least two distinct similarity-based models of inductive reasoning. The first, the similarity-coverage model, assumes that the probability that a blank property will be generalised from premise(s) to a conclusion depends both on the overlap between premise and conclusion categories (*similarity*) and the average maximum similarity of the premise categories to the lowest-level category that includes both the premises and the conclusion (*coverage*; Osherson et al., 1990). The similarity component is sufficient to explain premise-conclusion similarity effects, and the model accounts for premise typicality and premise diversity effects via increased levels of coverage from typical (higher mean similarity to conclusion category) and diverse premises (Hayes et al., 2010).

The second approach, the feature-based induction model, also assumes inductive generalisations depend on the similarity of premise and conclusion categories, but via a connectionist network that learns associations between premise and conclusion categories (Sloman, 1993). The model accounts for diversity effects in that training on a variety of mammals activates a greater number of connections than training on a narrower (non-diverse) range, strengthening any conclusions drawn.

These models have particular strength in explaining the use of inductive sampling heuristics that are based upon the taxonomic similarity between premise and conclusion categories. However, unlike the Relevance and Bayesian approaches discussed below, the models do not propose to account for the effects of causal knowledge on inductive generalisation.

Relevance theory. This model assumes that people evaluate inductive arguments by comparing premise and conclusion categories, activating the relationships between the categories that are most relevant for inductive generalisation (Medin et al., 2003). As causal relations are highly distinctive, they are more likely to be used as a basis for inference than taxonomic relations. When nonblank properties are used, these provide further clues as to which comparisons should be made between premise and conclusion categories (e.g., biological or behavioural). The diversity effect is accounted for by non-diverse premises activating a more distinctive property (e.g., ‘large cats’ in the case of lions and tigers) compared to diverse premises (e.g., polar bears and lions), which is less likely to be shared by the conclusion category (e.g., mammals).

Bayesian models. Under Bayesian models of induction (Heit, 1998; C. Kemp & Tenenbaum, 2009), people combine prior knowledge of category and feature relations with the information contained in argument premises in order to estimate the probability of a conclusion. C. Kemp and Tenenbaum's (2009) Bayesian model proposes that people have prior beliefs about structural relationships between different types of categories, and these relationships influence estimates of how properties are distributed. Different structures are accessed depending on the property to be generalised. For example, a ‘tree’ structure might capture taxonomic relationships between premise categories, with nearby objects in the tree (e.g., cows and horses) more likely to share both familiar and novel properties. Inferences about other kinds of categories or properties are thought to be represented using different structures (e.g., cause and effect networks when reasoning about disease transmission).

Are Deduction and Induction Driven by Separate Processes?

Traditionally, deduction and induction have been considered alternatives to each other, being studied separately with different paradigms, and with different theories postulated to account for reasoning performance (Heit, 2007). So far, the current thesis has followed this trend by discussing deduction and induction as if they are driven by separable processes. This issue, however, remains contentious, with an ongoing debate between single- and dual-process models of reasoning. This debate is now briefly reviewed.

Single-process models. Single-process accounts suggest that only one fundamental process underlies both deductive and inductive reasoning (Heit, 2007). In other words, there is essentially only one type of reasoning mechanism, and people apply this to problems that may be considered either deductive or inductive in nature (Harman, 1999).

Single-process accounts differ in their assumptions. Some accounts assume that people reason mainly in a probabilistic way: for example, Harman (1999) argues that people reason essentially non-deductively, whereas Oaksford and Chater (2007) argue that a Bayesian reasoning model can be applied to problems of both deduction and induction. Likewise, Osherson et al. (1990) and Sloman (1993) propose inductive models that can account for participant responses on particular types of deductive problems (e.g., arguments that include identical premises and conclusions are perfectly strong but are also deductively valid) without the addition of any further assumptions. Note that these accounts are not intended to account for all forms of deductive reasoning, such as syllogistic and conditional reasoning. They can, however, account for deductively valid arguments that can be assessed in terms of overlap between premise and conclusion categories.

A different single-process approach suggests that mental model theory (reviewed earlier in this chapter) can also be applied to problems of induction (Johnson-Laird, 1994). Here, the strength of any inference is assumed to depend on the relative proportion of possible states (represented by mental models) where the conclusion is true, relative to when it is false.

Dual-process models. Dual-process accounts, by way of contrast, suggest that there are two distinct kinds of reasoning that contribute to assessments of argument strength (Evans & Over, 1996; Sloman, 1996; Stanovich, 1999). It is assumed that one process is heuristic based, being relatively fast and influenced by context and associations, whereas the other is more deliberate, analytic, and rule-based (Darlow & Sloman, 2010; Evans, 2008; Heit, 2007; Stanovich, 1999). Inductive and deductive judgments are assumed to be influenced by both these processes, but to different degrees. It is thought that inductive judgments are particularly influenced by sampling heuristics (e.g., number of positive instances, diversity of premises) without necessarily being influenced by logical validity. In contrast, deduction is more likely to be deliberate and rule-based (Heit & Rotello, 2005; Heit et al., 2012).

Distinguishing between single- and dual-process accounts of reasoning. There is some neuropsychological evidence in favour of two separate reasoning systems. Even when reasoning about a common set of arguments, distinct brain areas are implicated when individuals are reasoning according to deduction instructions relative to induction instructions (Goel, Gold, Kapur, & Houle, 1997; Osherson et al., 1998). However, there have been relatively few attempts to experimentally compare induction and deduction. Rather, induction and deduction are typically studied separately (Heit, 2007), utilising tasks that differ across a variety of factors, such as task complexity and item content. This limits the scope to compare the reasoning processes underlying inductive and deductive tasks, often confounding possible reasoning differences with task differences.

One important exception was a study conducted by Rips (2001). He suggested that argument evaluation could be thought of as a rating on a single psychological dimension of argument strength, akin to judgments of loudness or brightness. This ‘criterion-shift’ account (Heit & Rotello, 2005) is illustrated in Figure 1.1. When judging inductive strength, one might set the criterion at Criterion 1 in Figure 1.1, representing the point that divides inductively strong (or plausible) arguments from those that are inductively weak (or implausible). Thus, Argument A represents an inductively strong argument. Deductively valid arguments, however, are only those with conclusions that necessarily follow from the given premises. Thus, when deciding where the separation between deductively valid (or correct) and deductively invalid (or incorrect) arguments should lie, the criterion must be shifted rightward (e.g., Criterion 2), where only the very strongest arguments can be called deductively correct (e.g., Argument B). This is the essence of the criterion-shift account; on this scale of argument strength, some arguments might be strong enough to be judged inductively plausible but not strong enough to be judged deductively correct, whereas some arguments are strong enough to satisfy the standard for deductive correctness.

Figure 1.1. has been removed due to Copyright restrictions. See Heit and Rotello (2005) for illustration of criterion-shift account of deduction and induction.

One prediction that follows from this account is that the relative order of arguments on this unitary scale of argument strength should remain the same, regardless of whether people are judging arguments deductively or inductively. Hence, arguments that are more likely to be rated as deductively valid should also be rated as inductively stronger than other arguments.

Rips (2001) tested this account by giving participants a common item set in which items varied in deductive validity and in their consistency with background causal knowledge. Notably, different groups were instructed to assess arguments in terms of either logical necessity (deduction condition) or in terms of plausibility (induction condition). Two argument types are of particular interest here: deductively valid but causally inconsistent arguments (e.g., ‘if Jill rolls in the mud, Jill gets clean; Jill rolls in the mud, therefore Jill gets clean’), and deductively invalid but causally consistent arguments (e.g., ‘Jill rolls in the mud, therefore Jill gets dirty’). In line with the ordinal predictions of the threshold account, participants given deduction instructions were more likely to endorse deductively valid but causally inconsistent arguments. However, those in the induction condition were more likely to endorse deductively invalid but causally consistent arguments (see Singmann & Klauer, 2011 for similar results). Rips (2001) considered this pattern a disconfirmation of the criterion-shift account.

Differential effects on inductive and deductive reasoning have also been demonstrated when items have been varied in other ways. For example, participants given deduction instructions are better able to discriminate between valid class-inclusion arguments (e.g., ‘all birds have Property X, therefore all sparrows have Property X’) and invalid arguments (e.g., ‘all sparrows have Property X, therefore all birds have Property X’), relative to those given induction instructions (Heit & Rotello, 2005, 2010; Rotello & Heit, 2009). In contrast, those given induction instructions are more sensitive to the inductive sampling heuristics of premise-conclusion similarity (Heit & Rotello, 2010) and argument length (Rotello & Heit, 2009) [2].

Implications of the single- vs dual-process debate for this thesis. Although the demonstration of dissociations between factors that affect induction and deduction are consistent with dual-process accounts, these demonstrations do not conclusively resolve the single- versus dual-process debate (for ongoing discussion see Evans & Over, 2013;

[2] It has been argued elsewhere that a single-process probabilistic model can account for ‘single dissociations’ between deduction and induction, where a factor influences one type of reasoning to a greater extent than the other (Lassiter & Goodman, 2015). However, this model cannot yet account for the ‘reversed association’ noted by Rips (2001), where those in the deduction condition were more likely to endorse valid/causally inconsistent arguments than invalid/causally consistent arguments, but this pattern reversed in the induction condition.

Lassiter & Goodman, 2015). More modestly, what this work *does* establish is that giving instructions to evaluate an argument either inductively or deductively causes people to focus on different features of otherwise identical arguments. These features include background knowledge and sampling heuristics (e.g., sample size) in the case of induction, and consistency with logical rules (item validity) in the case of deduction. For the purposes of this thesis, it is assumed that instructions to reason inductively or deductively will focus attention on different aspects of a given argument.

Moreover, it should be noted that the general method of presenting identical arguments for evaluation under either induction or deduction instructions (e.g., Heit & Rotello, 2010; Rips, 2001; Rotello & Heit, 2009) addresses a key weakness in much of the work comparing the two kinds of reasoning. That is, this method resolves issues of comparability between reasoning processes that are typically studied with tasks that differ on factors such as item content and task difficulty. Hence, this approach will be used extensively in the current experiments, which aimed to examine inductive and deductive performance in OCD[3].

Reasoning in Obsessive-Compulsive Disorder

To review, both deduction and induction are vital for everyday reasoning. Any impairment in deductive reasoning has flow-on implications for logic, scientific reasoning, and avoiding mistakes that follow invalid conclusions, whereas impairments in induction affect probabilistic reasoning, the expansion of knowledge, and decision-making. It is therefore important to explore how these processes manifest in OCD.

[3] Induction and deduction are not the only types of reasoning. Philosophers (e.g., Peirce, 1901) have also identified abduction (reasoning to the best explanation) as a component of human reasoning. For example, when told “a car stops suddenly”, one must then decide the likelihood that the car stops due to a particular cause (e.g., the car ran into a brick wall). Note that unlike induction and deduction, there have been very few studies about the psychological mechanisms underlying abduction (but see Lombrozo, 2010 for some recent relevant work). At present, therefore, it is difficult to carry out a systematic examination of the processes that underlie abduction in those with high or low OCD symptomatology.

It is clear from the earlier review of OCD symptomology that doubt and uncertainty over conclusions drawn from perceptual and behavioural evidence (e.g., seeing and knowing that one has turned off a light switch) is a key feature of OCD. Intuitively, the persistent urge to continue collecting uncertainty reducing evidence (Rotge et al., 2013; Stern et al., 2013) suggests that OCD patients may experience particular difficulty with reasoning that involves uncertain or probabilistic conclusions (i.e., inductive reasoning).

This view has been formalised in claims within the experimental psychopathology literature that OCD patients are unimpaired in their deductive reasoning abilities, but impaired when reasoning inductively (O'Connor, Aardema, & Pélissier, 2005; Pélissier & O'Connor, 2002). Henceforth, this will be referred to as the "spared deduction, impaired induction" account of reasoning in OCD. This relationship between deduction and induction is an important assumption underlying the Inference-Based Approach (IBA) model of OCD, which assumes that intrusions are inferences that have been formed through a process of impaired inductive reasoning (O'Connor et al., 2005).

Deductive Reasoning in Obsessive-Compulsive Disorder

To date, only two published studies have examined the deductive reasoning abilities of OCD patients. The first, a study by Pélissier and O'Connor (2002), administered three deductive tasks to OCD patients, generalised anxiety disorder (GAD) controls, and non-clinical controls. One task was the Wason Selection Task (Wason, 1966), which measures an individual's ability to use confirmatory and disconfirmatory evidence to validate a conditional logical rule. In this task, participants are presented with four cards, each with a letter on one side and a number on the other. Participants must turn over only the number of cards that will allow them to test a conditional rule, such as ‘if there is a vowel on one side of the card, there is an even number on the other side’. Typically, the task begins with four cards, one showing a vowel, one showing a consonant, one with an even number, and one with an odd number. The normatively correct response is to choose a pair of cards that can confirm and disconfirm the conditional rule (e.g., a card with a vowel and a card with an odd number). On this task, no differences were noted between OCD patient and control groups in the ability to use confirmatory or disconfirmatory evidence.

The second deductive task administered to OCD patients and controls by Pélissier and O’Connor (2002) involved the evaluation of logical syllogisms, in a task modelled after Evans, Barston and Pollard's (1983) procedure. Participants had to determine whether conclusions were justified solely by the evidence presented in the written passages. This is an example of an invalid/ believable argument (premises italicised, emphasis added): ‘Dogs are used extensively for the purpose of guarding property, guiding the blind and so on. *No highly trained dogs are vicious.* However, many people believe that their temperament cannot be trusted. The police service use dogs a great deal in their work. *Some police dogs are vicious* and although fatal accidents are rare, there is still growing concern over their widespread use. If the above passage is true, does it follow that: Some highly trained dogs are not police dogs?’[4]. Overall, there were no group differences in the number of ‘correct’ conclusions drawn (i.e., accepting valid conclusions and rejecting invalid conclusions), with healthy controls giving correct responses 71% of the time, and both OCD and GAD patients responding correctly 63% of the time.

[4] This example may not be identical to the arguments used by Pélissier and O'Connor (2002), who did not provide examples in their paper.

Lastly, Pélissier and O'Connor (2002) also administered a task adapted from the 2-4-6 problem (Wason, 1960), which required participants to produce numerical sequences in an attempt to replicate a rule only known by the experimenter (e.g., the rule behind the number sequence 2-4-6). Participants were asked to hypothesise aloud about the nature of the rule, with feedback given about the correctness or incorrectness of each hypothesis. There were no differences between groups in the number of individuals who discovered the rule, which was taken as further evidence that the deductive abilities of OCD patients are intact.

The second study to examine deductive reasoning in OCD also presented participants with a syllogistic reasoning task, with the two main differences being that these syllogisms were not embedded in passages and items were also varied according to OCD-relevance (Simpson, Cove, Fineberg, Msetfi, & Ball, 2007). The item set contained arguments that varied according to validity, believability and OCD-relevance. Valid arguments were those where the conclusion was necessitated by the premises, whereas for invalid arguments, the conclusion was merely possible, not completely necessary. Believability was also manipulated such that some conclusions were believable (e.g., ‘some rich people are not millionaires’) and some were unbelievable (e.g., ‘some millionaires are not rich people’). An example of a OCD-neutral valid/unbelievable syllogism was:

No well-educated people are fashionable

Some judges are fashionable

Some judges are not well-educated people

An example of an OCD-relevant invalid/believable syllogism was:

No life-threatening diseases are detectable

Some cancers are detectable

Some life-threatening diseases are not cancers

In Simpson et al.'s (2007) study, both the non-clinical control group and OCD patients were more likely to endorse valid conclusions and conclusions that were believable. In contrast to non-clinical controls, however, OCD patients exhibited poorer deductive performance on OCD-neutral items, judging invalid syllogisms as likely to be correct as valid ones. Both groups discriminated between valid and invalid items equally well when reasoning about OCD-relevant items. Simpson et al. (2007) argue that this different pattern of responding challenges the notion that there are no discrepancies between the deductive performance of OCD patients and controls. Rather, OCD patients seem to be deductively impaired when reasoning with OCD-neutral materials.

Inductive Reasoning in Obsessive-Compulsive Disorder

As first observed with clinical patients (O'Connor & Robillard, 1995, 1999), OCD patients seem to exhibit impairments in some forms of inductive generalisation. O'Connor and Robillard suggest that OCD patients generate distant possibilities on the basis of irrelevant associations, and then fixate upon these possibilities, rather than dismissing them on the basis of factual or observable evidence. As an example of this inductive generalisation style, O'Connor and Robillard (1995) cite a patient who believed that a table *must* be dirty (in spite of sensory evidence suggesting otherwise) because she saw a similarly shaped table a while ago that was dirty and as she had read in a magazine that tables easily accumulate dust.

There is some experimental evidence for deficits in the inductive reasoning abilities of OCD patients. Pélissier and O'Connor (2002) developed three tasks designed to compare the inductive performance of OCD patients to anxiety and non-clinical controls. The first task involved estimating the plausibility of 40 links between different premises and conclusions, whilst the second involved generating bridging statements between pairs of unrelated sentences (e.g., 'I went shopping downtown yesterday,' and 'the lady was very polite') as quickly as possible. The final task required participants to generate as many reasons as they could in support of the arbitrary statement 'this pen belongs to the experimenter' within a 2-minute time limit. Group differences were found in two of the three tasks. Firstly, OCD patients were slower to initiate and complete bridging statements than either of the control groups. Secondly, they were less convinced of the statement 'this pen belongs to the experimenter' *after* generating reasons for its support (whereas the conviction levels of the control groups remained essentially unmodified after the generation of supporting reasons). Decreases in conviction levels were positively related to the severity of OCD symptoms.

The authors hypothesised that these inductive differences were due to OCD patients considering alternative irrelevant ideas, which slowed down the generation of bridging statements, as well as leading to excessive doubting of the arbitrary statement. This notion was partly supported by the responses of the OCD patients, who alone chose to generate hypothetical reasons (e.g., 'maybe you were given the pen for your birthday') in favour of observable ones (e.g., 'the pen matches your outfit') in support of the arbitrary statement. These patterns of responding were taken to indicate that OCD patients exhibit a disordered form of inductive reasoning.

Aardema, O'Connor, Pélissier and Lavoie (2009) provided further support for the notion that OCD patients seem unduly influenced by irrelevant associations when reasoning inductively. They presented participants with scenarios about a possible accident (OCD-relevant) and a bus strike (OCD-neutral), with the information used to build the scenario alternating between being reality-based ('you look at people's faces and see no emotion that indicates an accident') and possibility-based ('the lack of expressions may have been shock'). Compared to non-clinical controls, OCD patients were more influenced by possibility-based rather than reality-based (observable) information in the OCD-relevant scenario, as defined by larger increases in probability ratings of whether the accident occurred following the presentation of possibility-based information. No group differences were noted in the OCD-neutral scenario.

OCD patients also place more emphasis on externally given rather than self-generated conclusions (as compared to non-clinical controls) when making inductive inferences (Pélissier, O'Connor, & Dupuis, 2009). Pélissier and colleagues presented participants with a range of OCD-neutral and OCD-relevant scenarios. An example of an OCD-neutral scenario was 'the man was bitten by a poisonous snake. There was no known antidote. What can you conclude?' Participants drew an initial conclusion and rated their confidence in their conclusion, before either generating three more conclusions ('self-generated' condition) or being presented with three additional conclusions ('given' condition). Participants then re-rated their confidence in their initial conclusion. There were no differences in initial levels of confidence between OCD patients and non-clinical controls. Unlike non-clinical controls, however, who were not sensitive to whether conclusions were self-generated or given, OCD individuals doubted the initial conclusion more when alternatives were given by the experimenter. Finally, OCD patients have also been noted to be less likely to propose definite solutions in an inductive 20-questions task (Simpson et al., 2007).

Limitations of Previous Clinical Research and Relation to Current Thesis

There are a number of limitations of the extant research examining deductive and inductive reasoning in OCD. This next section discusses these limitations and outlines how the current thesis sought to resolve these limitations.

Comparability of Deductive and Inductive Reasoning Tasks

One major difficulty with the current literature examining deductive and inductive reasoning in OCD is the same difficulty that exists in much of the wider reasoning literature: the tasks utilised thus far have not allowed for a direct comparison between the two reasoning processes, being confounded by task demands (e.g., difficulty and item content). For example, no two tasks in Pélissier and O'Connor's (2002) study involved remotely comparable content: in the deductive tasks, participants had to use given information to validate conditional rules or to decide whether conclusions were necessitated by premises, whereas in the inductive tasks, participants had to generate their own content as they attempted to link two unrelated sentences or justify an arbitrary statement.

Another difficulty is that the reasoning tasks employed have necessitated the integration of higher-order inductive and deductive processes with other mechanisms. For example, the generation of reasons in support of an arbitrary statement (Pélissier & O'Connor, 2002) not only involves induction, but the ability to construct and switch between a wide range of possible explanations. Hence, it is unclear whether the differences observed between OCD patients and controls on these tasks actually reflect impaired inductive reasoning or instead reflect difficulties with the additional task components.

In a similar vein, Wason's (1960) 2-4-6 rule discovery task, utilised as a deductive task in Pélissier and O'Connor's (2002) study, has previously been argued to involve inductive reasoning processes. Evans (1989), for example, suggests that the task involves the inductive generation of alternatives based on the given sequence and then a deductive assessment of whether these alternatives fit various possible underlying rules.

Thus, it is problematic to conclude that group differences in performance on inductive tasks are synonymous with differences in inductive ability on the basis of the extant evidence. It is equally difficult to conclude that a lack of group differences in performance on deductive tasks is equivalent to an absence of differences in deductive ability. Moreover, the many differences *between* the deductive and inductive tasks that have been employed thus far severely limits the ability to test the "spared deduction, impaired induction" account of OCD.

To enable a stronger test of this hypothesis, the experiments in the current thesis employed a number of strategies. Throughout this thesis, the examination of induction and deduction was primarily conducted with well-established measures of either induction (e.g., utilising inductive sampling heuristics) or deduction (e.g., manipulations of validity via class-inclusion); this reduces the likelihood that other processes besides induction and deduction contribute to task performance. As noted earlier, these reasoning tasks involved evaluating the validity or plausibility of arguments comprised of premise(s) and a conclusion (via binary responses and/or ratings). Unlike previous tasks used to examine deduction and induction in OCD (e.g., Aardema et al., 2009; Pélissier & O'Connor, 2002), there were no additional task requirements (e.g., to switch between alternative hypothetical scenarios).

Moreover, to enable a more direct comparison of deductive and inductive reasoning performance, many of the experiments in this thesis (Experiments 1, 2, 6 and 8) made use of the 'common stimulus set' paradigm developed by Rips (2001) and Heit and Rotello (2005, 2010, 2012). This involved presenting participants with a common set of arguments which are evaluated according to either inductive or deductive criteria.

The paradigm also involved manipulating factors thought to selectively influence one type of reasoning more than the other (e.g., logical validity for deduction; causal consistency for induction). This allowed for a direct comparison between deductive and inductive reasoning in individuals with differing levels of OCD symptomology, which was not confounded by task differences in stimulus content or complexity (Heit, 2007).

What Role Does General Ability Play in Deductive and Inductive Reasoning in Obsessive-Compulsive Disorder?

Normative performance on both deductive and inductive reasoning tasks has been shown to be positively related to general ability. For example, Stanovich and West (1998a) noted a moderate to large (r = .33-.50) positive correlation between correct deductive performance on syllogistic reasoning tasks (i.e., accepting logically valid arguments and rejecting invalid arguments) and a composite of cognitive ability (comprising Scholastic Aptitude Test scores, performance on the Ravens Matrices, and reading comprehension scores). A positive relationship was also noted between this composite and performance on the Wason Selection Task (r = .36). In terms of induction, increased use of the heuristics of monotonicity (r = .24) and diversity (r = .27) has been shown to be associated with higher levels of general ability, as indexed by the AH4 test of cognitive ability (Study 1; Feeney, 2007).

The association between general ability and normative inductive and deductive performance makes sense when one considers the processes involved in these tasks. Taking deductive reasoning as an example, experimental paradigms are often used which require participants to assume premises to be true solely for the purposes of mental simulation, even if the participants are uncertain about these premises (Evans, Handley, Neilens, & Over, 2007). This is a cognitively demanding process (Stanovich, 2011), and arguably, those with higher general ability will be better able to suppress at least some of their existing beliefs when reasoning deductively (Evans et al., 2007). Likewise, Kahneman and Frederick (2002) have argued that “intelligent people are more likely to possess the relevant logical rules and also to recognize the applicability of those rules in particular situations” (p. 68).

It is somewhat surprising then, that of the studies examining deductive and inductive reasoning in OCD, only Simpson et al. (2007) matched a non-clinical control group to the OCD patients on premorbid general ability and years educated. No other study examining deductive and/or inductive reasoning in OCD patients has taken into account the potential influence of general ability. Thus, the majority of experiments presented in the thesis included a measure of general ability, allowing for the contribution of general ability to reasoning performance to be examined.

General ability in obsessive-compulsive disorder. Previous studies that have compared the cognitive performance of OCD patients to that of non-clinical controls (but have not measured reasoning) have produced mixed results regarding general ability. The studies that indicate differences tend to find OCD patients impaired in the area of nonverbal ability (i.e., one's ability to think and organise material visually or spatially). For example, multiple studies have found that OCD patients perform poorly relative to non-clinical controls on the Performance subscales on the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), which involve such tasks as arranging blocks to match two-dimensional pictures (e.g., Boone, Ananth, Philpott, Kaur, & Djenderedjian, 1991; Bucci et al., 2007; Nakao et al., 2009). On tasks of verbal intelligence, as indexed by scores on the National Adult Reading Test (H. E. Nelson & Willison, 1991), and verbal subscales of Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981) and Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), OCD patients often do not differ from controls (e.g., Bannon et al., 2002; Bédard, Joyal, Godbout, & Chantal, 2009; Veale, Sahakian, Owen, & Marks, 1996).

With these considerations in mind, the majority of the present experiments included subtests of both verbal and non-verbal ability taken from the two-subtest version of the WASI. Although these subtests were not designed to be examined in isolation, the combined subtest scores allowed for an examination of whether differences in reasoning between those high or low in OCD symptoms exist over and above group differences in general ability.

Are Deductive and/or Inductive Differences Global or Specific?

Where deficits and/or biases in cognitive performance exist between psychiatric populations and controls, these differences may either be content-dependent (i.e., only appearing when threat-relevant material is involved) or content-independent (i.e., general differences that occur across a range of neutral and threat-relevant materials; McNally, 1999). If content-dependent, there are two lines of thought regarding how differences might manifest on threat-relevant material. Johnson-Laird, Mancini, and Gangemi (2006) suggest that the reasoning abilities of psychiatric populations should be *facilitated* due to their familiarity with emotionally-relevant content. In contrast, R. Kemp, Chua, McKenna, and David (1997) and McGuire, Junginger, Adams, Burright, and Donovick (2001) suggest that reasoning with threat-relevant materials in such populations should be *impaired*, possibly because of attentional biases toward threat-relevant material.

The one study that has indicated a deficit in deductive performance, with OCD patients being poorer at discriminating between valid and invalid syllogisms relative to controls, found this impairment on OCD-neutral and not OCD-relevant content (Simpson et al., 2007). In contrast, a review of tasks where inductive deficits have been noted in OCD patients have so far collectively indicated a global content-independent deficit, with differences present across tasks involving *both* OCD-neutral and OCD-relevant content. For example, Pélissier and O'Connor (2002) and Simpson et al. (2007) found impaired induction in OCD patients on tasks that contained content unrelated to patients' concerns. The one exception to the notion of global inductive impairment was the finding that OCD patients were more influenced by hypothetical possibilities rather than observable information when the scenario was OCD-relevant (Aardema et al., 2009).

Past comparisons of reasoning across OCD-neutral and OCD-relevant items are subject to many of the same methodological difficulties reviewed earlier (e.g., lack of comparability). Therefore, the majority of experiments in this thesis (Experiments 2-5 and 7-8) included both OCD-neutral and OCD-relevant stimuli, in order to more definitively determine whether deficits are global or specific to OCD-relevant content.

Aims of the Thesis

The overall aim of this thesis was to contribute to the literature regarding reasoning deficits in OCD. Specifically, this thesis aimed to:

1. Examine whether the "spared deduction, impaired induction" account of OCD reflects differences in the tasks used in previous clinical studies, rather than actual differences in deductive and inductive reasoning. Experiments 1, 2, 6 and 8 examined this issue using a well-established reasoning paradigm that resolves the methodological problems of previous studies by presenting participants with a common stimulus set, and only varying the instructions given (induction or deduction).

   Importantly, if individuals high in OCD symptoms are impaired inductively but exhibit intact deductive performance *on tasks where deduction and induction are directly comparable,* this would constitute additional novel evidence for a dual-process account of reasoning. That is, this pattern of responding would indicate that deduction and induction are driven by separate processes, with only one of these processes impaired in OCD.

2. Conduct a more exhaustive investigation of the nature of possible inductive deficits, by examining inductive reasoning involving inductive sampling heuristics, as well as inductive reasoning involving causal knowledge. Experiments examining the inductive heuristics of sample size (Experiment 1) and premise diversity (Experiments 2 to 5) are reported in Chapters 2 and 3. Inductive reasoning involving causal knowledge is examined in Experiment 6 (Chapter 4).
3. Determine whether any deductive and/or inductive biases or deficits are specific to OCD-relevant content or global across both OCD-neutral and OCD-relevant content. To that aim, Experiments 2-5 and 7-8 compared inductive and/or deductive reasoning performance on tasks that involved both OCD-neutral and OCD-relevant content.
4. Determine whether any group differences in deductive and/or inductive reasoning exist over and above any group differences in general ability (Experiments 2-8).

For the majority of experiments, analogue samples of university undergraduates (Experiments 1-4 and 6-7) or online participants (Experiment 5) who showed relatively high or low levels of OCD symptomatology were employed. This approach is justified for a number of reasons. First, analogue samples high in OCD symptoms have generally been noted to exhibit the same type of symptomatology as OCD patients, differing only by severity and the effectiveness of coping strategies utilised (Abramowitz et al., 2014; Gibbs, 1996). Indeed, high scorers on self-report measures of OCD symptoms often meet diagnostic criteria for OCD (Burns, Formea, Keortge, & Sternberger, 1995). Additionally, there is considerable evidence to indicate that the same developmental and maintenance factors (e.g., attentional biases, memory distrust, obsessive beliefs) are implicated in both clinical and analogue samples (Abramowitz et al., 2014), and there is also evidence of familial links between clinical and subthreshold forms of OCD (Gibbs, 1996). These factors have led Gibbs (1996) to suggest that the OCD symptomatology present in analogue samples might be considered a mild variant of OCD. Likewise, Abramowitz et al. (2014) conclude that, aside from severity, "OC symptoms in non-clinical individuals appear for the most part to be largely qualitatively indistinguishable from those in clinically diagnosed samples of OCD patients" (p. 213).

Importantly, in line with Abramowitz et al.'s (2014) recommendations, an empirically derived cut-off score was utilised to distinguish between high and low OCD symptom groups. Abramowitz et al. argues that this approach optimally identifies individuals who do and do not meet criteria for OCD, and is more likely to lead to generalisable results relative to arbitrarily derived cut-offs for a single sample (e.g., one standard deviation below or above the sample mean).

Where important differences between those high and low in OCD symptomatology were found, these differences were followed up in Experiment 8 with a *clinically diagnosed* sample of OCD patients. The concluding Chapter 6 summarises the key results of the empirical studies and discusses the theoretical, clinical, and methodological implications of the results.

CHAPTER 2 - Sensitivity to Premise Monotonicity and Argument Validity in Obsessive-Compulsive Disorder

Experiment 1

Previous accounts have suggested that OCD patients show deficits in inductive reasoning, but demonstrate relatively intact deductive reasoning abilities (e.g., Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). However, this “spared deduction, impaired induction” account of OCD has not yet been examined on tasks designed to directly compare deductive and inductive processes. Hence, a major aim of the current experiment was to examine reasoning performance in individuals with varying levels of OCD symptomatology on a task structured to directly compare deduction and induction.

The current procedure was patterned after that of Rotello and Heit (2009). Participants were asked to evaluate six types of arguments that varied in logical validity and sample size (premise monotonicity; see Table 2.1 for examples). Crucially, participants in different reasoning conditions were instructed to evaluate the same set of arguments on the basis of either logical necessity (deduction condition) or overall plausibility (induction condition). Because participants made inductive or deductive judgments about the same items, any obtained differences in reasoning performance could not be attributed to differences in item familiarity or difficulty.

Rotello and Heit (2009) noted that whilst the manipulation of argument validity had a larger impact on deductive than inductive judgments, manipulations of sample size primarily affected inductive judgments. This pattern fits with the conceptualisation of monotonicity as an inductive heuristic, where the strength or plausibility of an inductive conclusion tends to increase with the number of instances of positive evidence observed. This is because arguments based on many premises that share a property (e.g., the argument in the bottom right hand cell of Table 2.1) are seen as a better basis for property generalisation (or 'stronger') and more inclusive than arguments with fewer positive premises (e.g., the argument in the top right hand cell of Table 2.1; Osherson et al., 1990).

Table 2.1

Examples of the Six Argument Types Used in the Experiment 1 Premise Monotonicity Induction-Deduction Task

Argument length	Argument validity	
	Valid	Invalid
One-premise	All mammals have Property X ____________________ *All horses have Property X*	All donkeys have Property X ____________________ *All horses have Property X*
Three-premise	All mammals have Property X All wolves have Property X All goats have Property X ____________________ *All horses have Property X*	All donkeys have Property X All wolves have Property X All goats have Property X ____________________ *All horses have Property X*
Five-premise	All mammals have Property X All wolves have Property X All goats have Property X All monkeys have Property X All dogs have Property X ____________________ *All horses have Property X*	All donkeys have Property X All wolves have Property X All goats have Property X All monkeys have Property X All dogs have Property X ____________________ *All horses have Property X*

Note. The premises are given in normal font above the line and are assumed to be true. Conclusions are given in italics below the line. An argument is 'strong' if the conclusion is seen as plausible given the premises, and is 'valid' if the conclusion is seen as necessarily true.

Given that monotonicity effects (i.e., greater generalisation of a novel property based on an increasing number of positive premises) have been repeatedly demonstrated in studies with non-clinical subjects (e.g., Choi et al., 1997; Feeney, 2007; Heit & Rotello, 2012; McDonald et al., 1996; Osherson et al., 1990; Rotello & Heit, 2009), the extent to which an individual shows a monotonicity effect can be seen as one indicator of inductive functioning. Note that most previous work on the monotonicity effect has only used inductive arguments that are logically invalid (e.g., arguments like those in the right hand column of Table 2.1). Hence, like Rotello and Heit's (2009) study, it was expected that the manipulation of monotonicity would mainly impact the evaluation of invalid items under induction instructions in the current experiment.

In contrast, simply increasing the number of premises in an argument does not cause an argument to become 'more valid'. Thus, increasing argument length should not greatly increase argument acceptance under deduction instructions, particularly for logically valid items (Rotello & Heit, 2009). Indeed, somewhat surprisingly, Rotello and Heit found that increasing the number of premises in a logically valid argument led to a small but reliable *decrease* in argument acceptance under both inductive and deductive instructions. The authors argued that this may be due to valid arguments seeming more compelling (and therefore stronger) when they are briefer.

It was expected then, that individuals low in OCD symptoms in the current experiment would generally respond in a similar way to the undergraduates in Rotello and Heit's (2009) study. For invalid items, increasing the number of premises was expected to increase argument endorsement rates under induction instructions, but to have little effect on judgments of argument validity under deduction instructions. For valid items, increasing argument length was expected to decrease argument acceptance under both inductive and deductive instructions. In contrast, manipulation of argument validity was hypothesised to have a larger effect on argument evaluations in the deduction condition than in the induction condition. That is, those in the deduction condition were expected to exhibit greater discrimination between valid and invalid items, compared to those in the induction condition.

Individuals high in OCD symptoms were also expected to demonstrate the same sensitivity to argument validity, as these individuals are thought to exhibit intact deductive reasoning (O'Connor et al., 2005; Pélissier & O'Connor, 2002). However, given that the inductive reasoning abilities of individuals high in OCD symptomatology are thought to be impaired (Aardema et al., 2009; Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007), these individuals were not expected to demonstrate the same inductive sensitivity to increasing argument length as those low in OCD symptomatology.

The exact pattern of inductive performance for individuals high in OCD symptomatology was less easy to predict. As OCD patients often require more information than controls to make decisions or probability judgments (Fear & Healy, 1997; Milner et al., 1971; Volans, 1976), one possibility was that individuals high in OCD symptomatology would be less likely to accept invalid arguments with a given number of premises compared to those low in OCD symptomatology. Hence, they were expected to only show a weak or no monotonicity effect for invalid items under induction instructions. Similarly, these individuals were also expected to be less confident when making their inductive or deductive judgments.

This tendency to require greater amounts of information when making decisions or probability judgments raises a further possibility, however, that the extent to which an individual is *intolerant of uncertainty* may be more predictive of argument acceptance rates than an individual's level of OCD symptoms. Indeed, an increasing intolerance of uncertainty is associated with increasing levels of information search in ambiguous problem-solving tasks (Ladouceur, Talbot, & Dugas, 1997). Additionally, both individuals high in OCD symptoms and OCD patients with primary checking compulsions exhibit increased levels of intolerance of uncertainty relative to individuals low in OCD symptoms (Holaway, Heimberg, & Coles, 2006; Ladouceur et al., 1997). Therefore, another aim of the current experiment was to examine the extent to which OCD symptomatology affected rates of argument acceptance in comparison to levels of intolerance of uncertainty.

Method

Participants

One hundred and sixty-eight undergraduate students enrolled in an introductory first-year Psychology course at the University of New South Wales participated, after giving their informed consent. They were either recruited by telephone or via an online experiment sign-up system. All participants received course credit for their participation.

Participants were randomly assigned to either an induction or deduction condition. Three participants in the induction condition and one participant in the deduction condition were excluded from the analyses because they endorsed at least 97% of invalid items, leaving 82 participants in each reasoning condition.

Participants were also separated into groups based on their scores on the Obsessive-Compulsive Inventory-Revised (OCI-R; Foa et al., 2002). OCI-R items are found in Table 2.2. Individuals high in OCD symptoms were defined as those scoring equal to or greater than 21 on the OCI-R (n = 61, M = 30.75, SD = 8.25), consistent with the cut-off used in previous work for distinguishing between non-clinical controls and OCD patients (Foa et al., 2002). Of these participants, the majority were female (n = 49) and the mean age was 18.98 years (SD = 1.41). Low OCD symptom individuals were defined as those scoring lower than 21 on the OCI-R (n = 103, M = 10.93, SD = 5.41); the majority were female (n = 69) and the mean age was 19.85 years (SD = 5.14). There was no difference between the two OCD groups on age, F(1, 162) = 1.67, p = .20, but there was a trend toward there being a higher proportion of females in the high OCD symptom group, χ^2 (1, N=164) = 3.38, p = .07. Those high in levels of OCD symptoms were more intolerant of uncertainty (M = 36.41, SD = 9.70) than the low OCD symptom group (M = 24.50, SD = 8.58), F (1, 162) = 66.98, p < .001, η_p^2 = .29, as indexed by scores on the Intolerance of Uncertainty Scale (IUS, see Table 2.3; Carleton, Norton, & Asmundson, 2007).

Materials and Procedure

Each participant was tested individually in a quiet room. Participants began the experiment with a premise monotonicity argument evaluation task, before completing two questionnaires, the OCI-R and the IUS.

Premise monotonicity induction-deduction task. The argument evaluation task was created using the programming package LiveCode©. Participants were asked to make judgments about 60 arguments containing content from four categories: mammals, fruits, flowers and alcohol (see Appendix A for a complete list of stimuli). Each argument contained premises (or given statements) above a line and a conclusion statement below the line.

The arguments were divided equally by deductive validity: half were deductively valid and half were deductively invalid. Deductively valid arguments contained one premise of a category that included the conclusion premise in its scope. An example of a valid argument (i.e., an argument with a conclusion that is necessarily true, based on the given premises) is given in the top left-most cell of Table 2.1 (presented earlier). This is a valid argument because of the class-inclusion relation between horses and mammals.

Deductively invalid arguments were those where the conclusion was not necessarily true based on the given premises alone. An example of an invalid argument is given in the top right-most cell of Table 2.1: although all donkeys have Property X, it is not necessarily true that all horses also have Property X. There were two types of conclusions for invalid items: general conclusions about the superordinate category (e.g., ‘all mammals have Property X’) or specific conclusions about other category exemplars (e.g., ‘all horses have Property X’). Of the 30 invalid arguments, 12 featured general conclusions; one for each category (mammals, fruits, flowers, and alcohol) at each argument length level.

The arguments were also divided by argument length: one third of the arguments consisted of a single premise, whilst another third had three premises and the remaining third had five premises. Thus, there were six argument types in total (one-, three- and five-premise arguments at each level of validity). All properties were an uninformative or ‘blank’ combination of one randomly chosen letter followed by one randomly chosen number, designed to minimise the influence of background knowledge on judgments. Both the order of the premises and the presentation order of the items were randomised.

Reasoning condition differed only by the instructions given to the participants: those in the induction condition made judgments of argument strength (‘strong’/ ‘weak’), and those in the deduction condition made judgments of argument validity (‘valid’/ ‘invalid’). Following the instructions used by Rips (2001) and Rotello and Heit (2009), those in the induction condition were told that ‘strong’ arguments were those for which “the sentence below the line is *plausible*, given that the information above the line is true”, and ‘weak’ arguments were those for which “the sentence below the line is *implausible*, given that the information above the line is true”. Those in the deduction condition were told that ‘valid’ arguments were those for which “the sentence below the line is *necessarily true*, given that the information above the line is true", whereas 'invalid' arguments were those for which "the sentence below the line is *not necessarily* true, given that the information above the line is true".

Responses were made via mouse click. As can be seen in the example in Figure 2.1, each decision was followed by a confidence rating made via mouse click on a sliding scale from 1 (not confident at all) to 5 (very confident). Responses were not time limited, and each subsequent item was presented as soon as a response was made.

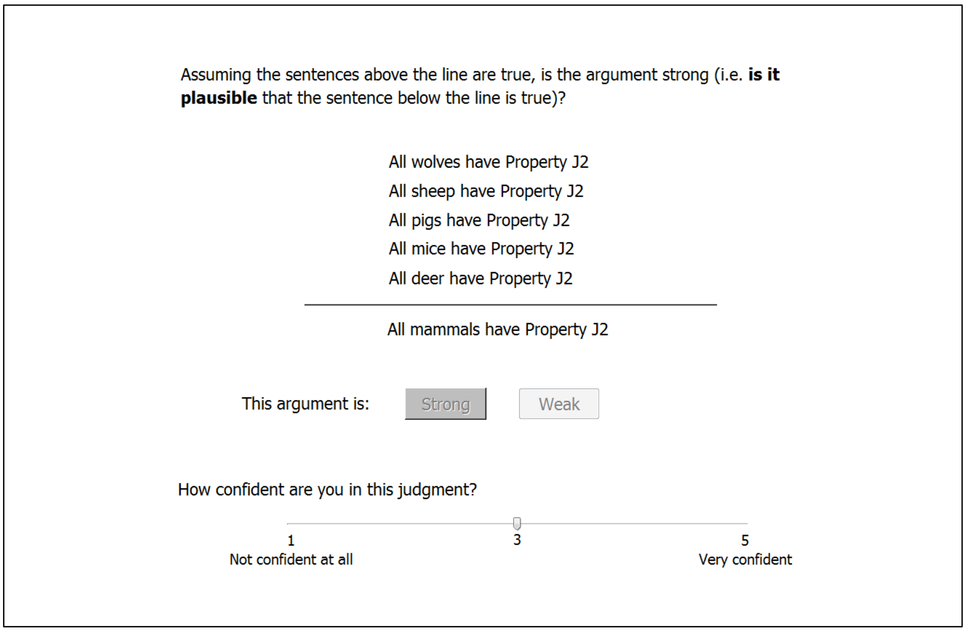

Figure 2.1. Screenshot of an inductive argument from the Experiment 1 premise monotonicity induction-deduction task that has been judged 'strong'.

Obsessive-Compulsive Inventory – Revised (OCI-R). The OCI-R (Foa et al., 2002) is an 18-item self-report scale that assesses the extent to which an individual experiences and is distressed by OCD symptomatology. As can be seen in Table 2.2, the scale covers six symptom categories (Washing, Obsessing, Hoarding, Ordering, Checking and Neutralising) commonly found in OCD, with three items assessing each category. Item ratings are made on a 5-point Likert scale (0 = not at all distressing, 4 = extremely distressing), and the final score is the sum of all 18 items (range = 0-72).

The OCI-R and its subscales have good internal consistency (α = .81 for OCD patients, α = .89 for non-clinical controls; Foa et al., 2002). The test-retest reliability for OCD patients over an approximately 2-week period is high (r = .82), as is the reliability for non-clinical controls over a 1-week period (r = .84). The OCI-R also has good convergent validity, and discriminates well between OCD and other anxiety disorders (Abramowitz, 2006).

Table 2.2. has been removed due to Copyright restrictions. See Foa et al. (2002) for 18-item Obsessive Compulsive Inventory-Revised (OCI-R).

Intolerance of Uncertainty Scale – Short (IUS). The IUS (Carleton et al., 2007) is a 12-item scale designed to measure the extent to which an individual considers the possibility of a negative event occurring unacceptable, irrespective of the likelihood of its occurrence (see Table 2.3 for IUS items). The IUS has high internal consistency ($\alpha = .91$) and is moderately correlated to measures of worry (Carleton et al., 2007). The test-retest reliability for non-clinical participants over an approximately 2-week period is satisfactory ($r = .77$), and scores on the IUS significantly predict pathological worry and trait anxiety in both clinically anxious and non-clinical samples (Khawaja & Yu, 2010). Further, individuals who are intolerant of uncertainty often interpret ambiguous information as threatening (Dugas et al., 2005) and a high intolerance of uncertainty may impair problem solving skills (Dugas, Freeston, & Ladouceur, 1997).

Table 2.3

The Intolerance of Uncertainty Scale – Short (IUS)

1. Unforeseen events upset me greatly.
2. It frustrates me not having all the information I need.
3. One should always look ahead so as to avoid surprises.
4. A small unforeseen event can spoil everything, even with the best of planning.
5. I always want to know what the future has in store for me.
6. I can’t stand being taken by surprise.
7. I should be able to organize everything in advance.
8. Uncertainty keeps me from living a full life.
9. When it’s time to act, uncertainty paralyses me.
10. When I am uncertain I can’t function very well.
11. The smallest doubt can stop me from acting.
12. I must get away from all uncertain situations.

Results

These analyses examined the proportion of positive responses; that is, the proportion of items that were evaluated as 'strong' (induction condition) or 'valid' (deduction condition) in binary judgments. Comparisons were drawn between individuals high and low in OCD symptoms. Relationships between scores on the OCI-R, IUS and the proportion of arguments accepted were explored through correlations and hierarchical regression analyses. Lastly, differences in confidence ratings were examined (see Appendix A for full analyses).

OCD is often conceptualised as a dimensional disorder (D. A. Clark, 2005; Mataix-Cols, Rosario-Campos, & Leckman, 2005). Arguably then, the highest scoring individuals on the OCI-R are those whose experience of OCD symptoms is closest to the experience of clinically diagnosed OCD patients. Thus, a strong test of whether OCD patients experience difficulties with inductive and/or deductive reasoning would be to examine performance differences between those scoring highest and lowest in OCD symptoms. For example, comparisons can be drawn between those scoring in the highest and lowest quartiles on the OCI-R (cf. Belayachi & Van der Linden, 2010; Myers & Wells, 2013; Ólafsson et al., 2014; Soref, Dar, Argov, & Meiran, 2008). Thus, in this experiment and subsequent experiments in this thesis, analyses were conducted on high and low OCD groups as defined both by a clinical cut-off and by those individuals showing the most extreme scores on the OCI-R (i.e., the lowest and highest quartiles).

Premise Monotonicity Induction-Deduction Task

Preliminary analyses. ANOVA analyses were conducted on the effect of conclusion type (general or specific) on the proportion of positive ('strong'/ 'valid') responses to invalid items. There was no main effect of conclusion type, $F(1, 160)$ =

0.34, p = .56, nor did conclusion type interact with any other factor (all F's < 2.41). Therefore, all further analyses were collapsed across conclusion type.

Proportion of positive responses. Initial analyses examined the proportion of positive ('strong'/ 'valid') responses to the six argument types (valid and invalid at each level of argument length). Means for valid and invalid items can be seen in Figure 2.2. The proportion of positive responses were analysed in a 2 (level of OCD symptoms: high or low) × 2 (reasoning condition: deduction or induction) × (2) (item validity: valid or invalid) × (3) (argument length: one-premise, three-premise or five-premise) ANOVA with repeated measures on the last two factors. Orthogonal linear and quadratic trend contrasts were planned in advance of the data to examine changes in evaluations of argument strength across levels of argument length.

The analysis revealed a main effect of validity, such that valid items were more likely to be accepted than invalid items, $F(1, 160) = 1120.75$, $p < .001$, $\eta_p^2 = .88$. Contrary to the expectation that those in the deduction condition would be more sensitive to item validity as compared to those in the induction condition, validity did not interact with reasoning condition, $F(1, 160) = 0.59$, $p = .45$. Validity did, however, interact with OCD group, $F(1, 159) = 17.52$, $p < .001$, $\eta_p^2 = .10$. Figure 2.3 shows that individuals high in OCD symptoms were *less* likely to accept valid arguments, but *more* likely to accept invalid items, compared to those low in OCD symptoms. That is, those high in OCD symptoms were less able to distinguish between arguments on the basis of logical validity than those low in OCD symptoms.

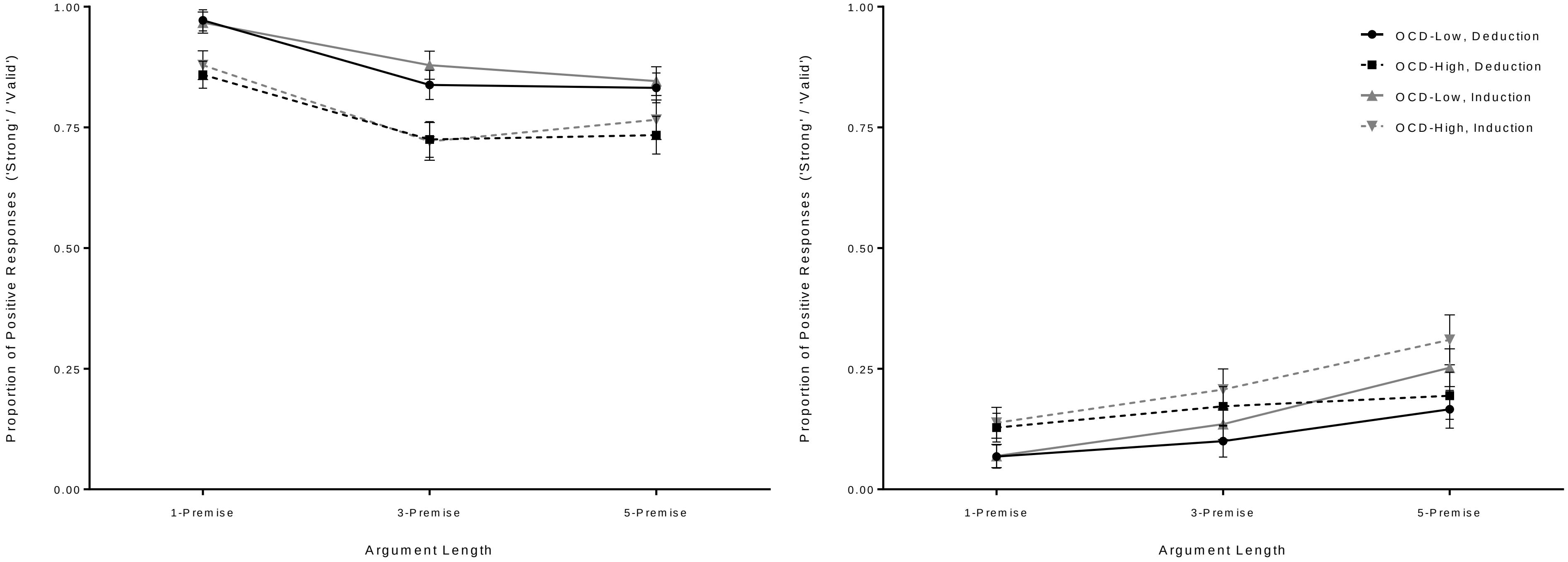

Figure 2.2. Proportion of positive responses (‘strong’/ ‘valid’) to valid (left panel) and invalid (right panel) items for three argument length levels by OCD group and reasoning condition on the Experiment 1 premise monotonicity induction-deduction task. Error bars represent standard error of the mean.

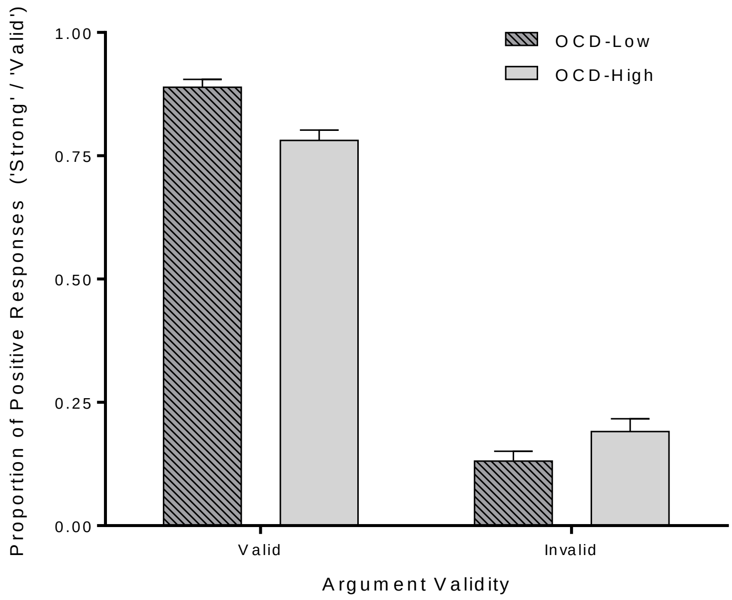

Figure 2.3. Proportion of positive responses (‘strong’/ ‘valid’) for valid and invalid items by OCD group on the Experiment 1 premise monotonicity induction-deduction task. Error bars represent standard error of the mean.

There was also a quadratic trend of argument length, such that one- and five-premise arguments were endorsed more often than three-premise arguments, $F(1, 160) = 22.68$, $p < .001$, $\eta_p^2 = .12$. As can be seen in Figure 2.2, this effect was qualified by an interaction with item validity, such that increasing argument length *increased* acceptance of invalid items, but *decreased* argument acceptance for valid items; linear trend, $F(1, 160) = 109.95$, $p < .001$, $\eta_p^2 = .41$, quadratic trend, $F(1, 160) = 13.23$, $p < .001$, $\eta_p^2 = .08$. Given that argument length seems to differentially affect valid and invalid items, responses to valid and invalid items are now considered separately.

Valid items. As can be seen in the left panel of Figure 2.2, when analyses were restricted to consideration of the valid items alone, increasing argument length led to decreased argument acceptance; linear trend, $F(1, 160) = 46.62$, $p < .001$, $\eta_p^2 = .23$, quadratic trend, $F(1, 160) = 31.42$, $p < .001$, $\eta_p^2 = .16$. This is consistent with the findings of Rotello and Heit (2009), who noted that valid arguments became weaker as they became longer. There was no main effect of reasoning condition, $F(1, 160) = 0.39$, $p = .53$, indicating that when responding to valid items, individuals did not differ in their responses despite being given different instructions. No other main or interaction effects reached significance.

Invalid items. As expected, for logically invalid arguments, increasing argument length led to a linear increase in argument acceptance, $F(1, 160) = 43.15$, $p < .001$, $\eta_p^2 = .21$. Moreover, this effect was stronger under induction than deduction instructions, $F(1, 160) = 5.82$, $p = .02$, $\eta_p^2 = .04$.

One of the main predictions was that under induction instructions, individuals high in OCD symptomatology would show less sensitivity to increasing argument length than individuals low in OCD symptomatology; i.e., they were expected to show impaired use of the monotonicity heuristic when reasoning inductively. To test this, OCD group comparisons were restricted to individuals in the induction condition. No differences in the effects of argument length on argument acceptance were found between those scoring high and low in OCD symptoms; linear trend, $F(1, 80) = 0.02$, $p = .89$, quadratic trend, $F(1, 80) = 0.07$, $p = .79$. Thus, contrary to expectations, there were no differences in the inductive performance of those high and low in OCD symptoms. No other main or interaction effects reached significance.

Lowest and highest quartiles of OCD-symptoms. Restricting comparisons to individuals with the highest and lowest OCD symptoms replicated all of the effects obtained in the earlier analyses (see Appendix A for full analyses). Those scoring highest in OCD symptoms were less likely to accept valid arguments but more likely to accept invalid arguments, relative to those scoring lowest in OCD symptoms, $F(1, 82) = 20.42$, $p < .001$, $\eta_p^2 = .20$. Once again, there was no significant interaction between argument length and OCD group; linear trend, $F(1, 82) = 0.49$, $p = .48$, quadratic trend, $F(1, 82) = 1.12$, $p = .29$.

Intolerance of uncertainty. One of the hypothesised variables that may affect the willingness to call an argument ‘strong’ or ‘valid’ is the extent to which an individual is intolerant of uncertainty. Scores on the OCI-R and IUS were highly positively correlated, indicating that the two scales shared a large proportion of variance, $r(162) = .60$, $p < .001$. To determine which of the two symptom measures (OCI-R or IUS) was more predictive of the ability to distinguish between valid and invalid arguments, hierarchical regression analyses (see Table 2.4) were conducted on difference scores (proportion of valid items endorsed – proportion of invalid items endorsed).

Table 2.4

Hierarchical Regression Analyses Predicting Ability to Distinguish Between Valid and Invalid Arguments on the Experiment 1 Premise Monotonicity Induction-Deduction Task by OCI-R and IUS According to Order of Entry

Step	Predictor variable	β	R^2	R^2 change	F change
Order of Entry 1: OCI-R entered first					
Step 1	OCI-R	-.34	.11	.11	20.52**
Step 2	OCI-R	-.38	.12	.003	0.60
	IUS	.07			
Order of Entry 2: IU entered first					
Step 1	IUS	-.15	.02	.02	3.95*
Step 2	IUS	.07	.12	.09	16.73**
	OCI-R	-.38			

Note. $N = 164$. β = standardised regression coefficient; OCI-R = Obsessive Compulsive Inventory-Revised; IUS = Intolerance of Uncertainty Scale. * $p < .05$, ** $p < .001$.

As can be seen in Table 2.4, OCI-R scores explained a significant amount of variance in sensitivity to validity, $p < .001$. The subsequent addition of the IUS factor did not lead to a significant increase in the amount of variance explained over and above that accounted for by OCI-R scores, $p = .44$.

When the order of entry of the variables was reversed, such that IUS was entered in Step 1, this variable accounted for a significant amount of variance in sensitivity to validity, $p = .05$. Notably however, when OCI-R scores were entered in Step 2, they accounted for a significant amount of additional variance in deductive performance, $p < .001$. This differential pattern indicates that whilst scores on both scales predicted the extent to which an individual differentiated between invalid and valid arguments, only the OCI-R accounted for additional unique variance over and above intolerance of uncertainty. This is a strong indication that OCI-R scores are a better predictor of reasoning performance than IUS scores[5].

Confidence ratings. The mean ratings of confidence in argument evaluations by reasoning condition and argument type are shown in Table 2.5.

Valid items. Participants were generally more confident when evaluating shorter than longer arguments; linear trend, $F(1, 160) = 33.34$, $p < .001$, $\eta_p^2 = .17$, quadratic trend, $F(1, 160) = 16.77$, $p < .001$, $\eta_p^2 = .10$. Those high in OCD symptoms were also generally less confident in their responses than those low in OCD symptoms, $F(1, 160) = 7.75$, $p = .006$, $\eta_p^2 = .05$. OCD group did not interact with reasoning condition, $F(1, 160) = 0.18$, $p = .67$, indicating that the reduced confidence of those high in OCD

[5] Given that high and low OCD groups did not differ in their sensitivity to premise monotonicity, both OCI-R and IUS scores were unlikely to predict any measure of sensitivity to argument length (e.g., responses to invalid five-premise arguments – responses to invalid one-premise arguments). Further regression analyses were conducted on the first experiment in which inductive group differences were noted (Experiment 3, Chapter 2), in order to examine the differential contribution of OCI-R and IUS scores to differences in inductive reasoning.

symptoms was true for both deduction and induction conditions. No other main effects or interactions reached significance.

Invalid items. Participants were once again more confident when evaluating shorter than longer arguments; linear trend, $F(1, 160) = 20.09$, $p < .001$, $\eta_p^2 = .11$. No other main effects or interactions reached significance.

Table 2.5

Mean Confidence Ratings by OCD Group, Reasoning Condition, Argument Validity, and Argument Length on the Experiment 1 Premise Monotonicity Induction-Deduction Task (Standard Deviations in Parentheses)

Argument validity	Argument length	Low OCD symptom group		High OCD symptom group	
		Deduction	Induction	Deduction	Induction
Valid	One-premise	4.56 (0.50)	4.60 (0.49)	4.38 (0.57)	4.30 (0.59)
	Three-premise	4.32 (0.60)	4.44 (0.51)	4.11 (0.54)	4.15 (0.69)
	Five-premise	4.31 (0.64)	4.41 (0.52)	4.11 (0.57)	4.20 (0.66)
Invalid	One-premise	3.95 (0.90)	4.08 (0.65)	3.80 (0.71)	3.78 (0.88)
	Three-premise	3.83 (0.94)	3.81 (0.74)	3.61 (0.74)	3.68 (0.78)
	Five-premise	3.81 (0.89)	3.68 (0.79)	3.62 (0.75)	3.54 (0.79)

Discussion

Previous work has suggested that OCD patients show a selective deficit in inductive reasoning, but an unimpaired ability to reason deductively (e.g., Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). The current experiment examined this hypothesis by comparing the reasoning performance of individuals with high and low levels of OCD symptoms on a paradigm that allowed for the direct comparison of deductive and inductive reasoning using a common stimulus set.

The results were somewhat surprising. Contrary to predictions, there was evidence that individuals high in OCD symptomatology were poorer at distinguishing between deductively valid and invalid arguments. Those high in OCD symptoms were less likely to endorse valid arguments, and more likely to endorse invalid items, compared to those low in OCD symptoms. These effects were found when OCD symptomatology was defined by: a) the suggested clinical cut-off in OCI-R scores, and b) the highest and lowest quartiles of OCI-R symptoms. Further, scores on a measure of OCD symptomatology were more predictive of the ability to distinguish between valid and invalid arguments, compared to scores on a measure of intolerance of uncertainty.

As expected from Rotello and Heit's (2009) study, argument length had a greater effect on evaluations of the argument strength of invalid items under induction than deduction instructions. Under induction instructions, invalid arguments with a larger number of confirming premises (e.g., all donkeys, wolves, goats, monkeys and dogs have ‘Property X’, therefore all horses have ‘Property X’) were seen as stronger than arguments with fewer premises (e.g., all donkeys have ‘Property X’, therefore all horses have ‘Property X’). This replicates the general effect of premise monotonicity consistently noted in other studies (e.g., Choi et al., 1997; Feeney, 2007; Heit & Rotello, 2012; McDonald et al., 1996; Osherson et al., 1990; Rotello & Heit, 2009). However, contrary to the predictions of the “spared deduction, impaired induction” account, both high and low OCD symptom groups used the number of positive premises as a guide to inductive plausibility. Thus, there was little evidence to suggest that individuals high in OCD symptomatology were impaired in their use of the inductive heuristic of monotonicity.

Increasing argument length also had a differing effect on invalid and valid arguments: for invalid arguments, argument acceptance increased with argument length, but for valid arguments, acceptance decreased with argument length. This provides additional evidence that participants use quantitatively different criteria to evaluate the logical validity and inductive plausibility of items. Decreasing acceptance of valid arguments with increasing argument length was also noted by Rotello and Heit (2009), who argued that valid arguments may seem more parsimonious (and therefore stronger) when they are simpler or briefer. That is, the addition of logically redundant premises may interfere with argument evaluation.

That the monotonicity effect did not differ across conclusion type (general or specific) differs from Feeney (2007), who found that monotonicity effects were stronger for arguments with general than specific conclusions. This may be explained by procedural differences between the experiments. For example, the monotonicity manipulation in Feeney’s study involved presenting participants with two or three-premise arguments, whereas the manipulation of monotonicity in the current experiment was across one-, three-, and five-premise arguments, and may therefore have been a stronger manipulation of argument length for both types of conclusions.

Lastly, individuals high in OCD symptomatology were less confident in their argument evaluations of valid items than were those with low symptomatology. This is consistent with the profile of OCD as being characterised by persistent doubt (Tolin et al., 2003).

The observed effect of OCD symptomatology on the evaluation of deductively valid and invalid items suggests that individuals high in OCD symptoms may be impaired when reasoning deductively. This is especially interesting in light of the notion that the class-inclusion relationships used to manipulate validity in this experiment were relatively straightforward. For example, by 8 years of age, most children already have a good understanding of the implications of this relationship (see Winer, 1980 for a review). Thus, those highest in OCD symptoms seem impaired on a fairly elementary form of deductive reasoning.

The finding of deductive impairment parallels Simpson et al.'s (2007) study, in which OCD patients exhibited poorer discrimination between valid and invalid syllogisms relative to non-clinical controls when item content was OCD-neutral. However, these findings run contrary to Pélissier and O'Connor's (2002) study, who noted no differences in the pattern of deductive judgments made by OCD patients, and anxiety and non-clinical controls on three deductive tasks. Note, however, that Pélissier and O'Connor used very different tasks from the current experiment to measure deductive performance (e.g., the Wason Selection Task). Elsewhere, it has been argued that the Wason Selection Task may not be the best method for assessing deductive performance since the deductively invalid strategy of searching for additional positive evidence can be seen as adaptive from the perspective of Bayesian inference (Oaksford & Chater, 1994). Hence, it could be argued that the current experiment provides a more direct test of the deductive abilities of individuals high in OCD symptoms.

Another possible explanation for the reduced sensitivity to argument validity noted in the high OCD symptom group is that the high and low OCD symptom groups differed in general ability. Recall that previous studies have sometimes found differences in general ability between OCD patients and controls (e.g., Boone et al., 1991; Bucci et al., 2007; Nakao et al., 2009). This is potentially important as both deductive and inductive performance is demonstrably related to general ability (e.g., Feeney, 2007; Stanovich & West, 1998a). Thus, the next experiment included a measure of general ability as a covariate when examining the effects of OCD symptomatology on deductive and inductive judgments.

The failure to find an effect of OCD status on the use of the inductive monotonicity heuristic was also somewhat surprising. This conflicts with previous research suggesting that OCD patients exhibit impaired performance on a range of inductive tasks (Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). Again, these discrepant results may reflect differences between the specific tasks used to study inductive reasoning in the respective experiments. The current task was designed to examine use of the inductive heuristic of premise monotonicity (i.e., when comparing inductive performance on one-, three- and five-premise items, the only factor that differed between the arguments was the number of premises each contained). In contrast, previous tasks labelled as inductive reasoning tasks may have involved additional processing components (e.g., switching between alternative hypothetical scenarios), and it is these components that may be impaired in OCD.

It is too early to conclude, however, that individuals high in OCD symptoms can readily use all inductive sampling heuristics. Inductive heuristics differ from each other in a number of ways, including difficulty, developmental trajectory and association with general ability. For example, some heuristics generalise information from single cases (e.g., premise similarity and typicality) and others generalise from multiple cases (e.g., monotonicity and diversity). The latter are likely to be more difficult to use, as they involve the more complex processing of combining information across multiple premises (Hayes, 2007).

The ability of those high in OCD symptoms to utilise the relatively complex heuristic of premise diversity is examined in the next chapter. The assumption underlying this choice is that if individuals high in OCD symptoms can utilise this heuristic (in addition to their already established ability to use sample size information), then these individuals should also be able to use simpler sampling heuristics that involve generalisation from a single case.

Finally, the current experiment did not address whether any impairment in the use of inductive sampling heuristics is specific to an individual's domain of concern. Previous work has suggested that impairments in inductive reasoning in individuals high in OCD symptomatology should be noted on OCD-neutral material (e.g., Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). However, it is possible that impairments (R. Kemp et al., 1997; McGuire et al., 2001) or advantages (Johnson-Laird et al., 2006) in the use of inductive sampling heuristics may be restricted to OCD-relevant materials. This possibility is examined in the next experiment.

CHAPTER 3 - Sensitivity to Premise Diversity and Argument Validity in Obsessive-Compulsive Disorder

The primary aim of this thesis was to examine the “spared deduction, impaired induction” account of OCD (e.g., O'Connor et al., 2005; Pélissier & O'Connor, 2002) using inductive and deductive tasks which are: a) more clearly grounded in previous theory and empirical work on each type of reasoning, and b) more comparable in that they differ only in the type of reasoning required. In Experiment 1, it was noted that individuals high in OCD symptoms seemed to be poorer at discriminating between valid and invalid arguments relative to those low in symptoms. These same high OCD symptom individuals were as able as low OCD symptom individuals to use sample size to guide their inductive judgments. That is, the deductive abilities of those high in OCD symptoms appeared to be impaired, whilst their inductive abilities were unimpaired.

Experiment 1 is only the second study to demonstrate deductive impairments in individuals high in OCD symptomatology (see Simpson et al., 2007). However, the contribution of general ability in accounting for this reduced sensitivity to argument validity has not yet been examined. This is important as higher general ability is related to an increased ability to make correct deductions (Stanovich, 1999; Stanovich & West, 1998a), apply logical rules (Kahneman & Frederick, 2002), and to assume premises to be true solely for the sake of mental simulation, as is necessary in deduction (Evans et al., 2007). Thus, an important aim of Experiment 2 was to examine whether OCD group differences in sensitivity to argument validity exist above any group differences in general ability.

The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was used to assess general ability in this and subsequent experiments. This scale was chosen for its brief administration time and its strong psychometric properties: the WASI has good convergent, discriminant, construct and clinical validity, high internal consistency, and high correlation with scores on the WAIS-III (J. C. Kaufman & Kaufman, 2001).

Another aim of the experiments in this chapter was to examine how individuals high in OCD symptoms apply the inductive heuristic of premise diversity to their judgments of argument plausibility. Whilst no differences were noted between those high and low in OCD symptoms in sensitivity to sample size (the monotonicity heuristic) in Experiment 1, this does not necessarily imply that individuals high in OCD symptoms routinely employ all inductive sampling heuristics.

Recall that, all things being equal, more dissimilar premise pairs within a superordinate category (e.g., cows and mice have ‘Property X’) are seen as stronger evidence for property generalisation (e.g., mammals have ‘Property X’) than less dissimilar pairs (e.g., cows and horses have ‘Property X’). Adult participants reliably select diverse arguments as representing stronger evidence for a particular conclusion over non-diverse arguments in forced-choice decisions (Osherson et al., 1990), even when reasoning diagnostically (Kim & Keil, 2003). Adults also more frequently choose diverse evidence over non-diverse evidence when hypothesis testing (López, 1995), and are more likely to generalise category labels to new exemplars after training with a diverse category, as compared to training with a more homogenous category (Hahn, Bailey, & Elvin, 2005). Furthermore, challenging children’s incorrect scientific beliefs is more effective if these beliefs are challenged on multiple diverse levels, rather than a single level (Hayes, Goodhew, Heit, & Gillan, 2003).

Both premise monotonicity and premise diversity are sampling heuristics that involve integration of information across multiple premises, and are arguably more complex than sampling heuristics that only involve comparison of single premises (e.g., premise similarity, premise typicality). This aligns with the developmental trajectories of these sampling heuristics, with sensitivity to premise diversity and monotonicity emerging relatively late in development (e.g., Guntheil & Gelman, 1997; López, Gelman, Guntheil, & Smith, 1992; Rhodes, Gelman, & Brickman, 2008). For example, López et al. (1992) found that 5- to 6-year olds judged arguments to be stronger when premises were typical and when they were similar to the conclusion category. That is, these children displayed sensitivity to the inductive sampling heuristics of premise typicality and similarity. In contrast, children did not begin to show an appreciation of diversity or monotonicity until 8 years of age, and only did so when generalising to a general category (i.e., generalising to all animals rather than a specific animal). Note that there are multiple hypothesised explanations for these developmental changes in inductive reasoning, such as unfamiliarity with reasoning material (Heit & Hahn, 2001; Shipley & Shepperson, 2006) or because children focus on the individual properties of each exemplar, rather than evaluating across multiple exemplars (Rhodes et al., 2008).

Increased use of monotonicity and diversity during inductive reasoning is also associated with higher levels of general ability in adults (Feeney, 2007). Presumably then, if individuals high in OCD symptoms are as sensitive to premise diversity as they are to premise monotonicity, they should also be well able to utilise single-case sampling heuristics, which are arguably simpler to employ. Given the relative complexity of the premise diversity heuristic and its widespread implications for categorisation and reasoning, the experiments in this chapter aimed to examine its use in individuals varying in OCD symptomatology.

A further innovation of the experiments in this chapter is that they examine inductive reasoning with items that vary in their relevance to major obsessional themes, such as safety and contamination. Where inductive differences have been previously found between OCD patients and controls, these have primarily been found either globally or on tasks that only utilised OCD-neutral item content (Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007), with one notable exception (see Aardema et al., 2009). This leads to the prediction that any differences that might be noted between high and low OCD symptom groups should appear on items with OCD-neutral content. However, it is also possible that group performance will differ on OCD-relevant content, perhaps with greater impairment due to the activation of underlying beliefs that interfere with reasoning (R. Kemp et al., 1997; McGuire et al., 2001). Alternatively, reasoning on OCD-relevant content may be inoculated, or even facilitated, by the increased familiarity that individuals high in OCD symptoms have with OCD-relevant material (Johnson-Laird et al., 2006). These alternatives were examined in the current series of experiments.

Experiment 2

Once again to facilitate the direct comparison of deductive and inductive reasoning, participants were asked to judge the argument strength of a common item set, in line with the methodology employed by Heit and Rotello (2010), Rips (2001), and Rotello and Heit (2009). Importantly, participants were given different instructions depending on reasoning condition; those in the deduction condition made judgments according to deductive validity, whereas those in the induction condition made judgments according to inductive plausibility.

No study has yet examined premise diversity under conditions where logical validity has also been manipulated. It was expected that reasoning condition would interact with argument type: that is, those in the deduction condition were expected to be more sensitive to changes in argument validity (Heit & Rotello, 2010; Rips, 2001; Rotello & Heit, 2009), regardless of the diversity of the premises (as this has no influence on the logical status of an item). In contrast, those given induction instructions were expected to be more sensitive to the inductive heuristic of premise diversity.

Table 3.1 illustrates the current design with arguments that vary on three factors: logical validity (class-inclusion), premise diversity, and OCD-relevance. It was expected that, in line with previous studies demonstrating that non-clinical populations are sensitive to evidence diversity (e.g., Heit, Hahn, & Feeney, 2005; López, 1995; Osherson et al., 1990), individuals low in OCD symptomatology would utilise diversity information in their inductive generalisations. That is, the inductive judgments of those low in OCD symptoms would be influenced by the diversity of the arguments (i.e., the arguments in the first and third rows would be seen as stronger than arguments in the second and fourth rows of Table 3.1). As increasing the diversity of premise pairs does not make an argument 'more valid', diversity was not expected to affect deductive judgments to the same extent. Moreover, for those low in OCD symptoms, deductive judgments were expected to be highly influenced by the validity of the argument (i.e., arguments in rows one and two of Table 3.1 would be more likely to be accepted than arguments in rows three and four).

The manipulation of argument validity offered a further opportunity to examine differences in deductive reasoning performance between individuals high and low in OCD symptoms. Under the assumption that the group differences in deduction found in Experiment 1 were robust, individuals high in OCD symptomatology were expected to exhibit poorer discrimination between valid and invalid arguments than those low in OCD symptomatology, and these differences were expected to persist when the effects of general ability were controlled.

Table 3.1
Examples of the Eight Argument Types Used in the Experiment 2 Premise Diversity Induction-Deduction Task

Argument type	OCD-relevance	
	OCD-neutral	OCD-relevant
Valid/ Diverse	All mammals have an ileal vein All mice have an ileal vein ________________ *All cows have an ileal vein*	All forms of money are contaminated by the bacteria hemonasella coli All dollar bills are contaminated by the bacteria hemonasella coli ____________________ *All gold coins are contaminated by the bacteria hemonasella coli*
Valid/ Non-Diverse	All mammals have an ileal vein All horses have an ileal vein ________________ *All horses have an ileal vein*	All forms of money are contaminated by the bacteria hemonasella coli All silver coins are contaminated by the bacteria hemonasella coli ____________________ *All gold coins are contaminated by the bacteria hemonasella coli*
Invalid/ Diverse	All cows have an ileal vein All mice have an ileal vein ________________ *All mammals have an ileal vein*	All gold coins are contaminated by the bacteria hemonasella coli All dollar bills are contaminated by the bacteria hemonasella coli ____________________ *All forms of money are contaminated by the bacteria hemonasella coli*
Invalid/ Non-Diverse	All cows have an ileal vein All horses have an ileal vein ________________ *All mammals have an ileal vein*	All gold coins are contaminated by the bacteria hemonasella coli All silver coins are contaminated by the bacteria hemonasella coli ____________________ *All forms of money are contaminated by the bacteria hemonasella coli*

Note. The premises are given in normal font above the line and are assumed to be true. Conclusions are given in italics below the line.

Further, if inductive reasoning is impaired in OCD (e.g., O'Connor et al., 2005; Pélissier & O'Connor, 2002), then individuals high in OCD symptoms should not exhibit the same sensitivity to premise diversity under induction instructions as those low in symptoms. This should be most apparent on invalid items, which have the greatest correspondence with items used in previous studies that have examined the diversity effect without manipulating argument validity. However, if those high in OCD symptomatology are not impaired in their use of inductive sampling heuristics (as in Experiment 1), then no differences should be noted between high and low OCD symptom groups in their preference for diverse evidence.

The use of OCD-neutral and OCD-relevant arguments allowed for an examination of whether any group differences in either inductive or deductive reasoning were domain-general or limited to content specifically related to OCD. Finally, in line with Experiment 1 and the pervasive role that doubt plays in the disorder, it was expected that those high in OCD symptoms would be less confident in making their judgments relative to those low in symptoms.

Method

Participants

One hundred and thirty undergraduate students enrolled in an introductory first-year Psychology course at the University of New South Wales participated in this experiment for course credit. Ten participants were excluded from the experiment because English was not their primary language.

Participants were randomly assigned to one of two reasoning conditions: induction or deduction. There were 60 participants in each reasoning condition. These participants were split into OCD symptomatology groups based on their scores on the OCI-R. Those with a score of 21 or above were defined as having high levels of OCD symptomatology (n = 42, M = 32.14, SD = 8.27). Of these, 28 were female, and the mean age was 20.83 years (SD = 3.66). Of the 78 participants in the low OCD symptom group (M = 10.60, SD = 5.98), 43 were female, and the mean age was 19.59 years (SD = 3.11). There was a trend toward the high OCD symptom group being older than the low OCD symptom group, $F(1, 114) = 3.84$, $p = .053$, $\eta_p^2 = .033$, but gender distribution did not differ across groups, χ^2 (1, N=116) = 1.07, p = .33.

Materials and Procedure

Participants completed the experiment individually in a quiet room, which began with the premise diversity induction-deduction reasoning task, followed by the OCI-R. To finish the experiment, all participants completed the WASI, administered by trained experimenters following the standard administration procedure set out in the manual.

Premise diversity induction-deduction reasoning task. The reasoning task was created using the programming package LiveCode©. Participants were asked to make judgments about 39 arguments in total: 36 of these were test items and the remaining items were attention-check items. Each item contained premises above a line and a conclusion below the line (see Table 3.1, presented earlier, for examples).

Half the test items were valid according to a simple deductive principle (class-inclusion) and half were invalid. Items were also equally divided by diversity status: half were non-diverse arguments (with two similar premises for invalid items, and a similar premise and conclusion for valid items), whilst the remaining items were diverse (with two dissimilar premises for invalid items, and a dissimilar premise and conclusion for valid items). Examples of diverse arguments can be seen in the first and third rows of Table 3.1 and non-diverse arguments can be seen in the second and fourth rows of Table 3.1 (see Appendix B for the full item set).

Assignment of items as diverse or non-diverse was based on preliminary testing of the similarity between premise categories from a wider range of stimuli, rated by an independent group of 19 participants who did not take part in the main experiment (see Appendix B for full similarity ratings). Similarity of premise content was assessed in two ways. Participants rated the similarity of the two premise categories in each item on a scale ranging from 1 (not at all similar) to 10 (very similar). Pairs of premise categories were also presented (e.g., 'cows and horses' vs. 'cows and mice') and participants were asked to choose the pair that they thought was more similar. These assessments were used to select premises for diverse and non-diverse OCD-neutral and OCD-relevant items. Of the items used in the current experiment, the OCD-neutral premise pairs that were nominated as diverse ($M = 3.91$, $SD = 1.59$) were reliably perceived as less similar than the non-diverse pairs ($M = 8.37$, $SD = 0.92$), $t(18) = 11.14$, $p < .001$. This was also true of items with OCD-relevant content; the premise pairs nominated as diverse ($M = 4.85$, $SD = 1.59$) were reliably perceived as less similar than the non-diverse pairs ($M = 8.27$, $SD = 0.80$), $t(18) = 10.81$, $p < .001$. In addition, the premise pairs nominated as non-diverse were always selected as more similar than those nominated as diverse in the binary choice task.

Premise and conclusion categories (e.g., cows, horses, mammals; gold coins, silver coins, money) were selected so that they would be relatively familiar to participants. However, the majority of properties attached to each (e.g., 'have an ileal vein', 'are contaminated by hemonasella coli') were invented 'blank' properties to avoid the activation of prior beliefs that may differentially affect ratings of argument strength (Osherson et al., 1990).

Every participant saw three sets of items with OCD-neutral content (related to mammals) and six OCD-relevant item sets (related to contamination and checking concerns). Each set contained four argument types (i.e., a valid/diverse item, a valid/non-diverse item, an invalid/diverse item, and an invalid/non-diverse item), with each item in a set paired with a different blank property (e.g., a different disease). Both the order of premises and the presentation order of the items were randomised for each participant.

Three attentional check questions were also included. These arguments were structured such that the conclusion category was completely unrelated to the premises and did not make logical sense (e.g., premises relating to animals and a conclusion relating to alcohol possessing biological properties). Participants were excluded from the experiment if they answered any of these questions incorrectly (i.e., called this argument 'strong' or 'valid' depending on condition).

As in Experiment 1, participants in the deduction condition were told that 'valid' arguments were those where the conclusion was *necessarily true* given the premises, and 'invalid' arguments were those where the conclusion was *not necessarily true*. Those in the induction condition were told that 'strong' arguments were those where the conclusion was *plausible* given the premises, and 'weak' arguments were *implausible*.

Participants saw example items before making binary decisions ('strong'/ 'weak' in the induction condition, 'valid'/ 'invalid' in the deduction condition) via mouse click on the computer program. Items remained on the screen for as long as it took participants to make a response. An example of item presentation can be seen in Figure 3.1. Following each decision, participants rated their confidence on a scale from 1 (not confident at all) to 5 (very confident).

Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). In all experiments that controlled for general ability, the two-subtest version of the WASI (incorporating Vocabulary and Matrix Reasoning) was used. The Vocabulary Scale contains 42 words and the Matrix Reasoning scale is comprised of 35 matrices. The WASI is normed for participants aged between 6 and 89 years.

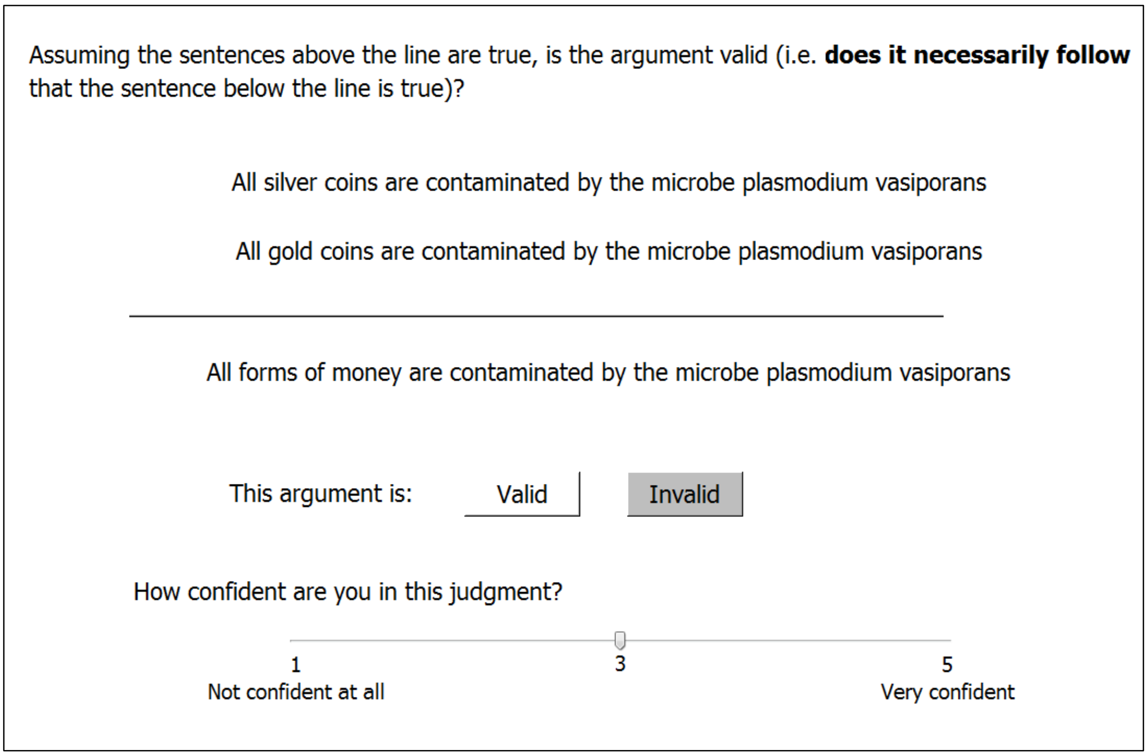

Figure 3.1. Screenshot of an invalid OCD-relevant item under deduction instructions on the Experiment 2 premise diversity induction-deduction task.

Results

These analyses examined the proportion of positive responses; that is, the proportion of arguments endorsed as ‘valid’ under deduction instructions and ‘strong’ under induction instructions in binary judgments. Comparisons were drawn between individuals high and low in OCD symptoms (OCI-R cut-off and quartile analyses). Following this, group differences in confidence ratings were examined (see Appendix B for full analyses).

Four participants were excluded from data analyses because they endorsed at least one of the attention check questions. Of the two that were in the induction condition, one was in the high OCD symptom group. The remaining two participants were in the deduction condition and were in the low OCD symptom group.

Preliminary Analyses

A one-way ANOVA analysis comparing high (*M* = 108.17, SD = 1.63) and low OCD symptom groups (*M* = 108.28, SD = 1.21) indicated that the two groups did not differ on estimates of general ability, $F(1, 114) = 0.00$, $p = .96$. This was also true of those scoring in the highest (*M* = 108.90, SD = 1.77) and lowest quartile of OCD symptoms (*M* = 111.03, SD = 1.63), $F(1, 59) = 0.76$, $p = .39$.[6]

Recall that general ability is associated with successful deductive reasoning and increased use of the diversity heuristic (Feeney, 2007; Stanovich & West, 1998a). General ability showed a small to moderate positive correlation with the ability to discriminate between valid and invalid arguments, as expressed by a difference score (proportion of valid items accepted - proportion of invalid items accepted); this was true for both OCD-neutral, $r(114) = .23$, $p = .01$, and OCD-relevant content, $r(114) = .37$, $p < .001$. However, no relationship was noted between general ability and the tendency to endorse invalid/diverse items relative to invalid/non-diverse items, for OCD-neutral, $r(114) = .09$, $p = .32$, or OCD-relevant items, $r(114) = .06$, $p = .55$. It is likely that the reduced number of items (e.g., three neutral items verses 48) obscured the usual relationship between general ability and diversity (e.g., Feeney, 2007). The relationship between premise diversity and general ability is revisited in Experiment 3.

---

[6] The two-subtest version of WASI only provides a full-scale estimate of general ability, and does not yield an estimate of verbal or non-verbal ability (Wechsler, 1999). As such, it is not appropriate to examine vocabulary or matrix reasoning performance separately. However, given that general ability differences between individuals with and without OCD have been noted primarily in the area of non-verbal ability (e.g., Boone et al., 1991; Bucci et al., 2007; Nakao et al., 2009), where differences between high and low OCD groups in general ability were noted, between OCD group repeated measures analyses were conducted with repeats on normed vocabulary and matrix reasoning scores. All these analyses indicated that general ability differences were consistent across both subtests (all *F*'s < 1.70).

Although the relationship between general ability and reasoning performance is a theoretically interesting issue, the primary focus of the current thesis is the relationship between OCD symptoms and reasoning. The relationship between general ability and the ability to discriminate between valid and invalid arguments illustrates the importance of controlling for general ability when examining other differences between OCD groups. Therefore, in the current experiment and all subsequent experiments where general ability was measured, these scores were used as a covariate. Accordingly, all means reported have been adjusted for general ability (i.e., estimated marginal means are reported). Where OCD group differences in reasoning performance change when general ability is not controlled, this is reported separately in a footnote.

Premise Diversity Induction-Deduction Task

Proportion of positive responses. Initial analyses examined the proportion of positive (‘strong’/ ‘valid’) responses to the eight argument types (valid/diverse, valid/non-diverse, invalid/diverse, and invalid/non-diverse with both OCD-neutral and OCD-relevant content). Estimated marginal means can be seen in Figure 3.2. The proportion of positive responses was analysed in a 2 (level of OCD-symptoms: high or low) × 2 (reasoning condition: deduction or induction) × (2) (item-validity: valid or invalid) × (2) (item-diversity: diverse or non-diverse) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA with repeated measures on the last three factors and controlling for general ability.

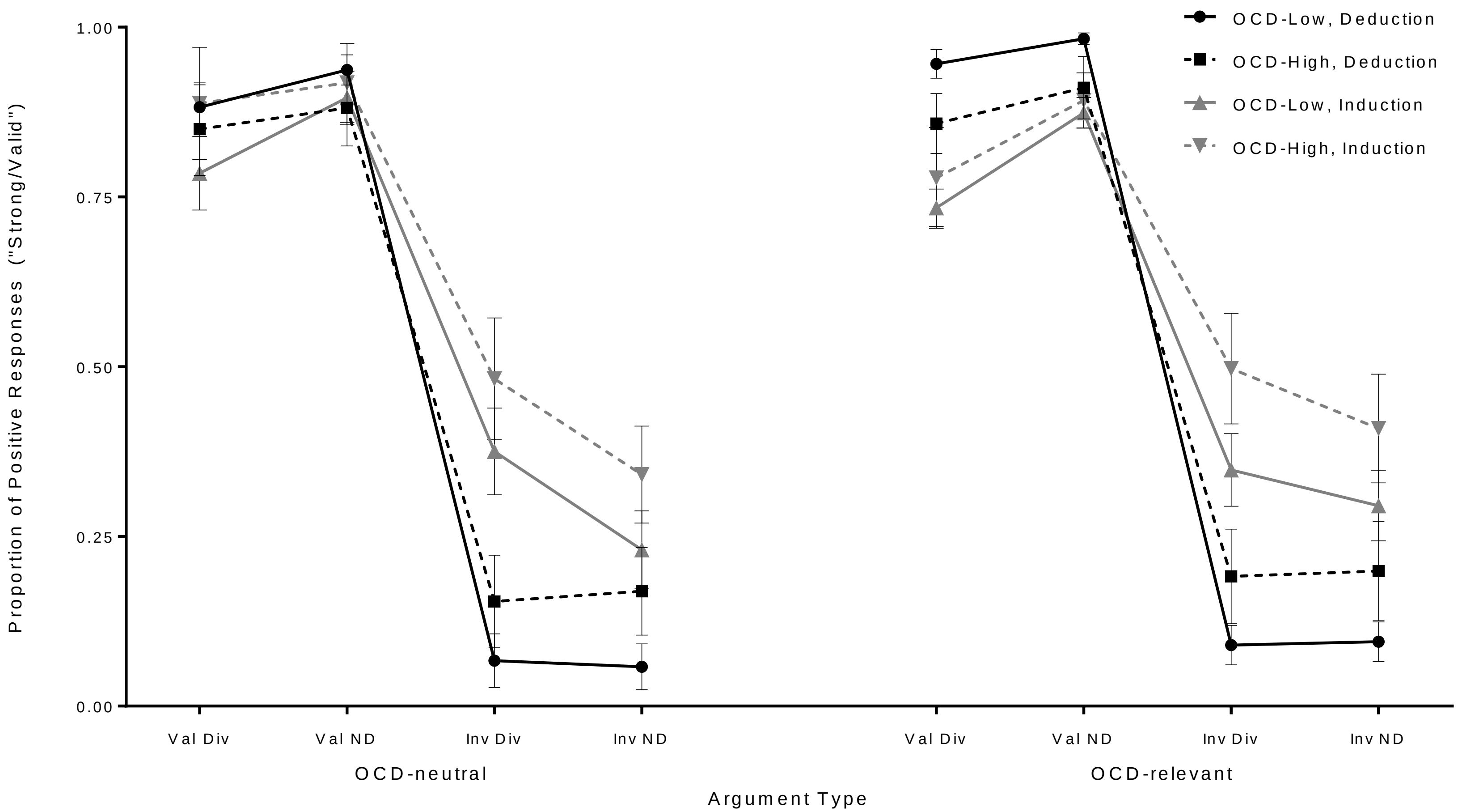

Figure 3.2. Estimated marginal mean proportion of positive responses ('strong'/ 'valid') by OCD-relevance, OCD group, reasoning condition, and argument type on the Experiment 2 premise diversity induction-deduction task. Val Div = Valid/Diverse; Val ND = Valid/Non-Diverse; Inv Div = Invalid/Diverse; Inv ND = Invalid/Non-Diverse. Error bars represent standard error of the mean.

Analyses revealed a main effect of reasoning condition, such that those given induction instructions accepted more arguments than those given deduction instructions, $F(1, 111) = 14.14$, $p < .001$, $\eta_p^2 = .11$. There was also a main effect of OCD-relevance, such that arguments were more likely to be accepted when they contained OCD-relevant content, $F(1, 112) = 11.37$, $p = .001$, $\eta_p^2 = .09$. Finally, there was also a main effect of validity, with valid arguments accepted more frequently than invalid arguments, $F(1, 112) = 359.88$, $p < .001$, $\eta_p^2 = .76$. No other main effects were significant.

As can be seen in Figure 3.2, validity interacted with reasoning condition, such that participants given deduction instructions were more sensitive to the validity of an argument than those given induction instructions, $F(1, 112) = 21.36$, $p < .001$, $\eta_p^2 = .16$. In line with Rips (2001) and Rotello and Heit (2009), those in the deduction condition showed greater discrimination between valid and invalid arguments, being more willing to accept valid arguments and less willing to accept invalid arguments, compared to those in the induction condition.

There was also an interaction between validity and diversity, $F(1, 112) = 20.78$, $p < .001$, $\eta_p^2 = .16$, such that when items were valid, a greater number of non-diverse arguments were accepted relative to diverse arguments. In contrast, when items were invalid, a greater number of diverse arguments were accepted relative to non-diverse arguments.

Finally, there was a three-way interaction between reasoning condition, validity, and diversity, $F(1, 112) = 9.55$, $p = .003$, $\eta_p^2 = .08$. This can be explained by differences in the interaction between reasoning condition and diversity-status for valid and invalid arguments. As can be seen in Figure 3.3, for valid items, the diversity status of the arguments did not differentially affect argument acceptance under deduction and induction conditions, $F(1, 112) = 3.54$, $p = .06$. In contrast, for invalid items, those in the induction but not the deduction condition were more likely to accept diverse arguments relative to non-diverse arguments, $F(1, 112) = 6.70$, $p = .01$, $\eta_p^2 = .06$. No other main effects or interactions reached significance.

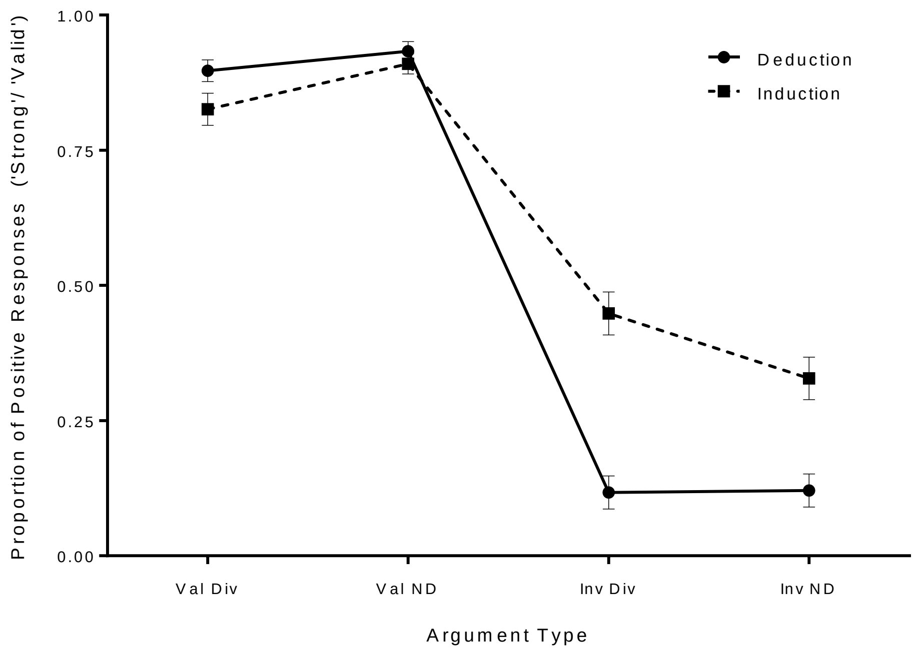

Figure 3.3. Estimated marginal mean proportion of positive responses (‘strong’/ ‘valid’) by reasoning condition and argument type on the Experiment 2 premise diversity induction-deduction task. Val Div = Valid/Diverse; Val ND = Valid/Non-Diverse; Inv Div = Invalid/Diverse; Inv ND = Invalid/Non-Diverse. Error bars represent standard error of the mean.

Given the results of Experiment 1, it was expected that individuals high in OCD symptoms would be impaired when reasoning deductively with class-inclusion items. Validity showed a marginal relationship with OCD group, such that individuals high in OCD symptoms trended toward being less able to distinguish between valid and invalid arguments relative to those low in OCD symptoms, $F(1, 112) = 3.47$, $p = .07$, $\eta_p^2 = .03$. However, when comparisons were restricted to only individuals scoring in the top and bottom quartiles of OCI-R scores, this difference reached significance, $F(1, 57) = 4.41$,

p = .04, η_p^2 =.05. That is, those highest in OCD symptoms were less able to distinguish between valid and invalid arguments relative to those lowest in OCD symptoms. Notably, these differences remained after the effects of general ability had been controlled.

One of the main predictions of this experiment was that under induction instructions, individuals high in OCD symptoms would show less sensitivity than individuals low in OCD symptoms to the premise diversity of invalid arguments. However, in the induction condition, no interaction between item diversity and OCD group was found on invalid items, $F(1, 56) = 0.81$, $p = .37$. This indicates that those high and low in OCD symptoms did not differ in their general preference for diverse over non-diverse invalid items. This was also true when the analysis was restricted to those scoring in the upper and lower quartiles of OCD symptoms, $F(1, 28) = 0.32$, p = .58.

Note, however, that the current task required binary judgments of argument strength for each argument presented. Thus, participants could accept (judge as ‘strong’) or reject (judge as ‘weak’) as many arguments as they liked. This differs from the method used to examine diversity in most previous studies, where participants made forced-choice decisions between diverse and non-diverse arguments to examine their preference for diversity (e.g., Kim & Keil, 2003; López et al., 1992; Osherson et al., 1990; Rhodes et al., 2008).

Hence, to better approximate the forced-choice nature of previous tasks, for each instance (i.e., set) of item content, comparisons were restricted to invalid items on which individual participants accepted a diverse argument but not the non-diverse argument (or vice versa). For example, if a participant judged all mammals to have an ‘ileal vein’ because all cows and horses have an ‘ileal vein’ (non-diverse premise pair), but rejected this same argument when told that all cows and mice have an ‘ileal vein’ (diverse pair), this item was included in the analysis as a preference for non-diversity.

Table 3.2 shows the overall proportion of trials where the same response was given to the diverse and non-diverse options, where different responses were given and the diverse option was labelled strong, and where different responses were given and the non-diverse option was labelled strong. An ANCOVA analysis comparing groups high and low in OCD symptoms whilst controlling for general ability was conducted on the proportion of instances where the non-diverse option was preferred (i.e., the non-diverse option was judged to be ‘strong’ and the diverse option was judged to be ‘weak’) when different judgments were made on the diverse and non-diverse versions of an item. The proportion of such ‘non-diverse’ responses was low overall (<19% responses). Nevertheless, for OCD-neutral items, those high in OCD symptoms were more likely to endorse non-diverse as opposed to diverse arguments relative to those low in OCD symptoms, $F(1, 55) = 6.63$, $p = .013$, $\eta_p^2 = .11$. This was not the case for OCD-relevant content, where the preference for non-diverse items did not differ between those high and low in OCD symptoms, $F(1, 55) = 0.19$, $p = .92$.

Table 3.2

Proportion of Trials on which Same or Differing Responses were Given to Diverse and Non-diverse Options by OCD Group and OCD Relevance on the Experiment 2 Diversity Induction-Deduction Task (Standard Deviations in Parentheses).

Proportion of trials	Low OCD group		High OCD group	
	Relevant	Neutral	Relevant	Neutral
Same response given	0.68 (0.24)	0.72 (0.33)	0.74 (0.25)	0.56 (0.35)
‘Diverse’ labelled strong	0.23 (0.20)	0.21 (0.33)	0.17 (0.23)	0.25 (0.31)
‘Non-diverse’ labelled strong	0.09 (0.14)	0.07 (0.14)	0.09 (0.13)	0.19 (0.23)

Confidence ratings. The estimated marginal mean confidence ratings for each reasoning condition, OCD group, and argument type are shown in Table 3.3. In contrast to Experiment 1, there was a main effect of reasoning condition, $F(1, 111) = 6.74$, $p = .01$, $\eta_p^2 = .06$, and no main effect of OCD group, $F(1, 111) = 2.69$, $p = .10$. That is, participants in the deduction condition were more confident in their judgments than those in the induction condition, but individuals high and low in OCD symptoms were equally confident when making their judgments.

As in Experiment 1, participants were more confident making judgments regarding valid ($M = 4.35$, $SD = 0.59$) rather than invalid ($M = 3.99$, $SD = 0.72$) arguments, $F(1, 112) = 29.98$, $p < .001$, $\eta_p^2 = .21$. Validity and reasoning condition interacted, such that the reduced confidence of those in the induction condition (valid, $M = 4.29$, $SD = 0.59$; invalid, $M = 3.77$, $SD = 0.72$) relative to the deduction condition (valid, $M = 4.41$, $SD = 0.60$; invalid, $M = 4.22$, $SD = 0.64$) was more pronounced on invalid items than valid, $F(1, 112) = 5.02$, $p = .03$, $\eta_p^2 = .04$. There was a main effect of diversity status, such that participants were more confident making judgments involving non-diverse ($M = 4.22$, $SD = 0.58$) rather than diverse ($M = 4.12$, $SD = 0.61$) items, $F(1, 112) = 10.83$, $p < .001$, $\eta_p^2 = .09$. Diversity also interacted with validity, $F(1, 112) = 18.37$, $p < .001$, $\eta_p^2 = .14$, such that the increased confidence participants felt when judging non-diverse arguments was larger for valid items (Valid/Diverse, $M = 4.25$, $SD = 0.67$; Valid/Non-Diverse, $M = 4.45$, $SD = 0.58$; Invalid/Diverse, $M = 4.00$, $SD = 0.72$; Invalid/Non-Diverse, $M = 3.99$, $SD = 0.77$).

Table 3.3

Estimated Marginal Mean Confidence Ratings (out of 5) by OCD-Relevance, OCD Group, Reasoning Condition, Argument Validity, and Argument Diversity on the Experiment 2 Premise Diversity Induction-Deduction Task (Standard Deviations in Parentheses).

OCD-		Low OCD group		High OCD group	
relevance	Argument type	Deduction	Induction	Deduction	Induction
Neutral	Valid/Diverse	4.17 (0.85)	4.27 (0.77)	4.19 (1.15)	3.80 (0.87)
	Valid/Non-Diverse	4.44 (0.77)	4.49 (0.68)	4.47 (0.62)	4.10 (0.72)
	Invalid/Diverse	4.21 (0.76)	3.80 (0.91)	4.10 (0.60)	3.73 (0.93)
	Invalid/Non-Diverse	4.24 (0.82)	3.95 (0.88)	4.27 (0.70)	3.62 (1.00)
Relevant	Valid/Diverse	4.45 (0.56)	4.43 (0.59)	4.30 (0.67)	4.03 (0.65)
	Valid/Non-Diverse	4.52 (0.53)	4.63 (0.50)	4.42 (0.67)	4.22 (0.44)
	Invalid/Diverse	4.24 (0.70)	3.82 (0.66)	4.03 (0.76)	3.95 (0.70)
	Invalid/Non-Diverse	4.18 (0.75)	3.75 (0.73)	4.12 (0.61)	3.69 (0.76)

Additionally, OCD-relevance interacted with validity, F (1, 112) = 6.92, p = .01, η_p^2 = .06, such that the participants' increased confidence when judging valid items over invalid items was larger on OCD-relevant content relative to OCD-neutral content. Finally, OCD-relevance also interacted with diversity, F(1, 112) = 6.34, p = .01, η_p^2 = .05; the increased confidence participants felt when judging non-diverse items was larger when items were OCD-neutral.

Collectively, the pattern of confidence ratings seems to suggest that participants are more confident when judging arguments that involve certainty (i.e., deductive judgments, valid arguments) or lower levels of dissimilarity (i.e., non-diverse premises).

Further, these results suggest that individuals high in OCD symptoms are not always less confident when reasoning than those low in OCD symptoms.

Discussion

Previous work has suggested that OCD patients have a selective deficit in inductive reasoning but an unimpaired ability to reason deductively (e.g., Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). The current experiment re-examined this hypothesis by comparing the inductive and deductive performance of individuals high and low in OCD symptomatology on a common stimulus set.

In line with Experiment 1, there was evidence to suggest that individuals highest in OCD symptoms were poorer at distinguishing between deductively valid and invalid arguments, relative to those lowest in symptoms. That is, those scoring in the highest quartile of OCD symptoms accepted fewer valid arguments and a greater number of invalid arguments relative to those scoring in the lowest quartile of OCD symptoms. Importantly, these differences were found after controlling for general ability, indicating that differences in sensitivity to validity are not explained by group differences in general ability. Thus, those highest in OCD symptoms do seem impaired on an elementary form of deductive reasoning.

As expected, under induction instructions, arguments based on diverse premises were generally regarded as stronger (or more convincing) than those based on non-diverse arguments, replicating the established inductive heuristic of premise diversity (e.g., Kim & Keil, 2003; López, 1995; Osherson et al., 1990). No differences were noted between OCD groups in the ability to utilise the diversity heuristic on invalid items, with both groups accepting more diverse than non-diverse premises. Ostensibly, this could be taken to indicate that individuals high in OCD symptomatology are as able to use the diversity heuristic as those low in OCD symptomatology.

However, this method represents a conservative test of sensitivity to diversity, because for each instance of item content, both the diverse and non-diverse options could be accepted, or both could be rejected. When comparisons were restricted to occasions where participants accepted only one of these options (i.e., either accepted the invalid/diverse argument and rejected the invalid/non-diverse argument, or vice versa), individuals high in OCD symptoms were more likely than individuals low in OCD symptoms to have accepted the non-diverse argument on OCD-neutral items. This seems to indicate that there may be OCD group differences in sensitivity to the diversity heuristic. However, given that this difference was only found on a subset of items, Experiment 3 employed a less conservative test of diversity by presenting participants with forced-choice decisions regarding argument strength between diverse and non-diverse evidence pairs.

Experiment 3

To further explore the finding in Experiment 2 suggesting that individuals high in OCD symptoms prefer non-diverse arguments when reasoning about OCD-neutral item content, participants were administered a more direct test of sensitivity to premise diversity. This task required participants to choose whether arguments based on diverse or non-diverse premises represented stronger evidence for inductive generalisation. Further, participants were asked to rate the inductive strength of each argument. As the results of Experiment 2 indicated that inductive sensitivity to premise diversity may be reduced in OCD, it was expected that individuals high in OCD symptoms would show a reduced preference for diverse evidence, both in forced-choice decisions and in ratings of inductive strength, relative to those low in symptoms.

Method

Participants

One hundred and eleven undergraduate students enrolled in an introductory first-year Psychology course at the University of New South Wales participated. All received course credit for their participation. In line with Experiment 2, 11 participants were excluded from the experiment as English was not their primary language.

Once again, for the purpose of analysis, participants were separated into groups based on their scores on the OCI-R (Foa et al., 2002). Forty-four participants were defined as high OCD symptom participants, scoring equal to or greater than 21 on the OCI-R (M = 29.95, SD = 7.19); of these participants, the majority were female (n = 30) and the mean age was 18.75 years (SD = 1.64). There were 56 low OCD symptom individuals (M = 10.86, SD = 5.42); the majority of these participants were female (n = 30) and the mean age was 20.46 years (SD = 6.67). The two groups did not differ in age, $F(1, 98) = 2.77$, $p = .10$, nor in the distribution of gender across groups, $\chi^2(1, N=100) = 0.14$, $p = .16$.

Materials and Procedure

Participants completed the experiment individually in a quiet testing room. All participants completed two reasoning tasks: the forced-choice diversity task and a different induction-deduction task, which is described in a later chapter (see Experiment 6, Chapter 4). Task order was counterbalanced across participants. All participants then completed a computerised version of the OCI-R and IU. Finally, the two-subtest version of the WASI was administered by the author in accordance with the procedures set out in the manual.

Forced-choice premise diversity reasoning task. This inductive reasoning task was created using the programming package Livecode©. Within the program, participants were presented with 30 arguments, each composed of two pairs of premises (one diverse, one non-diverse) and a general conclusion (see Appendix B for the full set of stimuli). One premise was common to both premise pairs, and the other premise in each pair was designed to be either similar (non-diverse) or dissimilar (diverse) from the shared premise. Half of the items presented were OCD-neutral (related to mammals), whilst the other half were OCD-relevant (containing emotional content related to common OCD-relevant concerns, such as washing and checking). Examples of these items can be seen in Rows 3 and 4 of Table 3.1 (that is, the structure of these items was the same as the invalid items used in Experiment 2).

Assignment of diverse or non-diverse content was again based on pre-test ratings of similarity between premise content drawn from the same wider range of stimuli as in Experiment 2 (see Appendix B for full similarity ratings). Pre-testing established that, for this reduced item set, the OCD-neutral premise pairs that were nominated as diverse (M = 3.80, SD = 1.27) were reliably perceived as less similar than the non-diverse pairs (M = 7.16, SD = 0.63), $t(18) = 14.75$, $p < .001$. This was also true of OCD-relevant items; the premise pairs nominated as diverse (M = 5.31, SD = 1.50) were reliably perceived as less similar than the non-diverse pairs (M = 7.92, SD = 0.68), $t(18) = 10.51$, $p < .001$.

Participants were asked to make forced-choice selections between two sets of premise pairs (diverse or non-diverse), indicating which set was more likely to convince them that a particular ‘claim’ (or conclusion) was true. Responses were made via mouse click. An example of an item can be seen in Figure 3.4. Arguments were not labelled according to diversity-status. The left-right positioning of diverse and non-diverse pairs was randomised, as was item order. After each forced-choice selection, participants rated the usefulness of the diverse and non-diverse arguments in judging whether the conclusion category was true on a scale of 1 (not useful at all) to 10 (very useful).

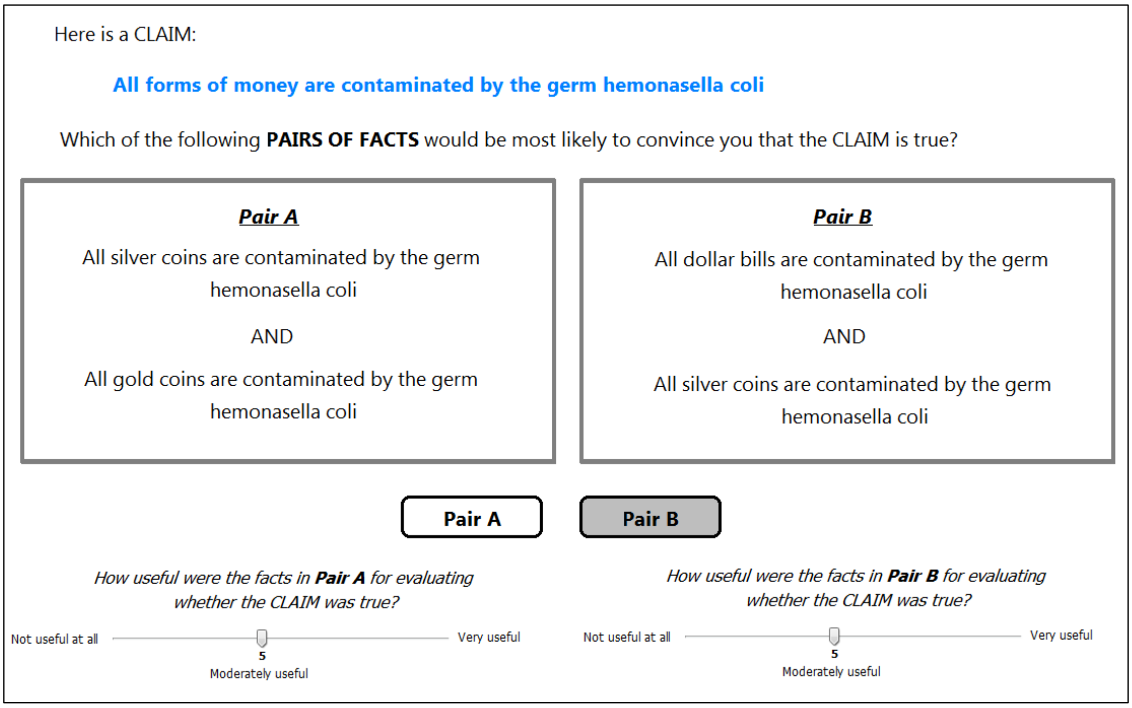

Figure 3.4. Screenshot of an item from the Experiment 3 forced-choice premise diversity program with the diverse pair selected.

Results

Analyses were conducted on the proportion of diverse arguments selected as providing stronger evidence for a particular conclusion. Comparisons were conducted between individuals high and low on OCD symptoms (clinical cut-off and quartile analyses). Following this, OCD group differences in ratings of evidence strength (usefulness) for each type of argument (diverse or non-diverse) were examined (see Appendix B for full analyses).

Preliminary Analyses

ANOVA analyses were conducted on task order (premise diversity task first or induction-deduction task first), demonstrating no systematic effects (all *F*'s < 2.72) on the proportion of diverse arguments selected. Therefore, all further analyses were collapsed across task order.

A one-way ANOVA analysis of general ability scores between high and low OCD symptom groups indicated that the groups did not differ in estimates of general ability, $F(1, 98) = 3.36$, $p = .07$. However, general ability estimates for those scoring in the highest quartile on the OCI-R ($M = 105.37$, $SD = 11.62$) were lower than those scoring in the lowest quartile ($M = 112.42$, $SD = 7.55$), $F(1, 49) = 6.41$, $p = .02$, $\eta_p^2 = .12$. Further, general ability scores exhibited a small to medium positive relationship with the proportion of diverse premise pairs endorsed, $r(98) = .21$, $p = .04$. As in the previous experiment, full-scale general ability scores were included as a covariate in subsequent analyses of reasoning performance.

Forced-Choice Premise Diversity Task

Proportion of diverse arguments endorsed. The proportion of diverse arguments endorsed was examined in a 2 (level of OCD-symptoms: high or low) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA with repeated measures on the last factor and general ability included as a covariate. As can be seen in Table 3.4, overall, both high and low OCD symptom groups were more likely to choose the diverse arguments as providing stronger evidence for inductive generalisation than would be expected by chance, for both OCD-neutral items and OCD-relevant items. Table 3.4 shows a trend towards a weaker diversity effect in the high OCD symptom group relative to the low OCD symptom group, but the main effect of OCD group did not reach significance, $F(1, 97) = 2.69$, $p = .10$.

Table 3.4

Estimated Marginal Mean Proportion of Diverse Arguments Endorsed by OCD Group and OCD-Relevance on the Experiment 3 Forced-Choice Premise Diversity Task (Standard Deviations in Parentheses) and Details of Significance Tests Compared to Chance Responding (0.50)

	Low OCD group				High OCD group			
Sample	Proportion diverse	*t*	*df*	*p*	Proportion diverse	*t*	*df*	*p*
OCD-neutral items								
Clinical cut-off	0.76 (0.30)	6.86	55	<.001**	0.67 (0.31)	3.41	43	.001*
Quartile	0.85 (0.33)	8.79	23	<.001**	0.61 (0.29)	1.69	26	.103
OCD-relevant items								
Clinical cut-off	0.75 (0.22)	8.62	55	<.001**	0.68 (0.24)	4.56	43	<.001**
Quartile	0.81 (0.24)	9.83	23	<.001**	0.63 (0.24)	2.52	26	.018*

Note. $^*p < .05$, $^{**}p < .01$.

However, when OCD group comparisons were restricted to those individuals showing the most extreme scores on the OCI-R (i.e., the highest and lowest quartiles), a reliable difference was found in the strength of the diversity effect between OCD quartile groups. The group highest in OCD symptoms showed weaker sensitivity to diversity than the group lowest in OCD symptoms, $F(1, 48) = 10.09$, $p = .003$, $\eta_p^2 = .17$. OCD quartile group did not interact with OCD-relevance, $F(1, 49) = 2.04$, $p = .16$, indicating that the reduced preference for diverse evidence demonstrated by the highest OCD symptom group relative to the lowest OCD symptom group did not differ across OCD-neutral and OCD-relevant items. For OCD-neutral items, the proportion of diverse responses endorsed by those with the highest level of OCD symptoms did not differ from chance. Both quartile groups preferred the diverse option when item content was OCD-relevant (see Appendix B for full quartile analyses).

Intolerance of uncertainty. Given that inductive differences were noted between high and low OCD quartile groups, the contribution of IUS to this group difference was examined. Scores on the OCI-R and IUS were highly positively correlated, $r(98) = .74$, $p < .001$; this relationship was stronger than in Experiment 1. Checks for multicollinearity noted no problems. To determine which of the two variables was more predictive of sensitivity to argument diversity, hierarchical regression analyses were conducted on the proportion of diverse options endorsed (see Table 3.5). For each step, the contribution of OCI-R and IUS scores over and above the contribution of general ability scores to sensitivity to argument diversity was examined.

OCI-R scores explained a significant amount of variance in sensitivity to argument diversity controlling for general ability, $p = .008$. The subsequent addition of the IUS factor in Step 2 did not increase the amount of variance explained, over and above that already accounted for by general ability and OCI-R, $p = .72$. When the order of entry of the variables was reversed, such that IUS was entered before OCI-R, IUS scores accounted for a significant amount of variance in sensitivity to diversity, $p = .03$. When OCI-R scores were entered in Step 2, they did not account for a significant amount of additional variance in inductive performance, $p = .12$.

Thus, neither OCI-R nor IUS scores accounted for additional variance in inductive sensitivity when entered in Step 2 of the model. This is unsurprising for two reasons: first, OCI-R and IU scores were more highly correlated in the current experiment than they were in Experiment 1. Second, an intolerance of uncertainty is likely to be highly relevant to reasoning with uncertain and highly probabilistic inductive conclusions. Overall, OCI-R scores appear to be the better predictor of deductive reasoning (Experiment 1) and do not differ from IUS in the ability to explain inductive reasoning performance. Therefore, the OCI-R was used as the primary group measure throughout the remainder of the thesis.

Table 3.5

Hierarchical Regression Analyses Predicting Proportion of Diverse Arguments Endorsed on the Experiment 3 Forced-Choice Premise Diversity Task by OCI-R and IUS According to Order of Entry

Step	Predictor variable	β	R^2	R^2 change	F change
Order of Entry 1: OCI-R entered second					
Step 1	IQ	.14	.11	.07	7.32**
	OCI-R	-.27			
Step 2	IQ	.14	.11	.001	.717
	OCI-R	-.23			
	IUS	-.05			
Order of Entry 2: IU entered second					
Step 1	IQ	.17	.09	.05	4.87*
	IUS	-.22			
Step 2	IQ	.14	.11	.02	2.45
	IUS	-.05			
	OCI-R	-.23			

Note. N = 100. β = standardised regression coefficient; IQ = Pro-rated general ability estimates on the Wechsler Abbreviated Scale of Intelligence; OCI-R = Obsessive Compulsive Inventory-Revised; IUS = Intolerance of Uncertainty Scale.
* p < .05, ** p < .001.

Ratings of perceived usefulness. These data were analysed in a 2 (level of OCD-symptoms: high or low) × (2) (item diversity: diverse or non-diverse) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA, with repeated measures on the last two factors. The estimated marginal mean usefulness ratings for diverse and non-diverse arguments are given in Table 3.6.

Table 3.6

Estimated Marginal Mean Perceived Usefulness Ratings (out of 10) by OCD-Group and OCD-Relevance on the Experiment 3 Forced-Choice Premise Diversity Task (Standard Deviations in Parentheses)

	Low OCD group		High OCD group	
Sample	Diverse	Non-diverse	Diverse	Non-diverse
OCD-neutral items				
Clinical cut-off	5.89 (1.83)	4.72 (1.68)	5.55 (1.73)	4.93 (1.53)
Quartile	5.85 (1.73)	4.31 (1.37)	5.38 (1.67)	5.03 (1.41)
OCD-relevant items				
Clinical cut-off	6.00 (1.71)	4.92 (1.59)	5.55 (1.63)	4.85 (1.49)
Quartile	5.92 (1.68)	4.53 (1.64)	5.33 (1.31)	4.85 (1.45)

Overall, participants rated arguments based on diverse premises as more useful than those based on non-diverse premises, $F(1, 98) = 34.05$, $p < .001$, $\eta_p^2 = .26$. There was no main effect of OCD-relevance, $F(1, 98) = 1.27$, $p = .26$, or OCD group, $F(1, 97) = 0.29$, $p = .59$; this indicates that there were no overall differences in perceived usefulness ratings for OCD-neutral or OCD-relevant items, nor between groups high and low in OCD symptoms. There was also no interaction between OCD group and diversity, $F(1, 98) = 2.32$, $p = .13$, indicating that those high in OCD symptoms did not differ from those low in OCD symptoms on the usefulness ratings they assigned to diverse and non-diverse arguments.

However, when the analysis was restricted to the highest and lowest quartile of OCD scorers, there was an interaction between OCD group and item diversity, $F(1, 49) = 8.76$, $p = .005$, $\eta_p^2 = .15$. This can be seen in Figure 3.5. Bonferroni-adjusted simple effects analyses with an alpha level of .0125 indicated that those lowest in OCD symptoms assigned higher usefulness ratings to arguments based on diverse premise pairs relative to the non-diverse pairs, $F(1, 49) = 32.19$, $p < .001$, $\eta_p^2 = .40$, while those highest in OCD symptoms did not differ in the usefulness ratings they assigned to diverse and non-diverse premises, $F(1, 49) = 2.90$, $p = .10$. There were no OCD quartile differences in usefulness ratings assigned to either diverse, $F(1, 48) = 0.89$, $p = .35$, or non-diverse arguments, $F(1, 48) = 1.36$, $p = .25$.

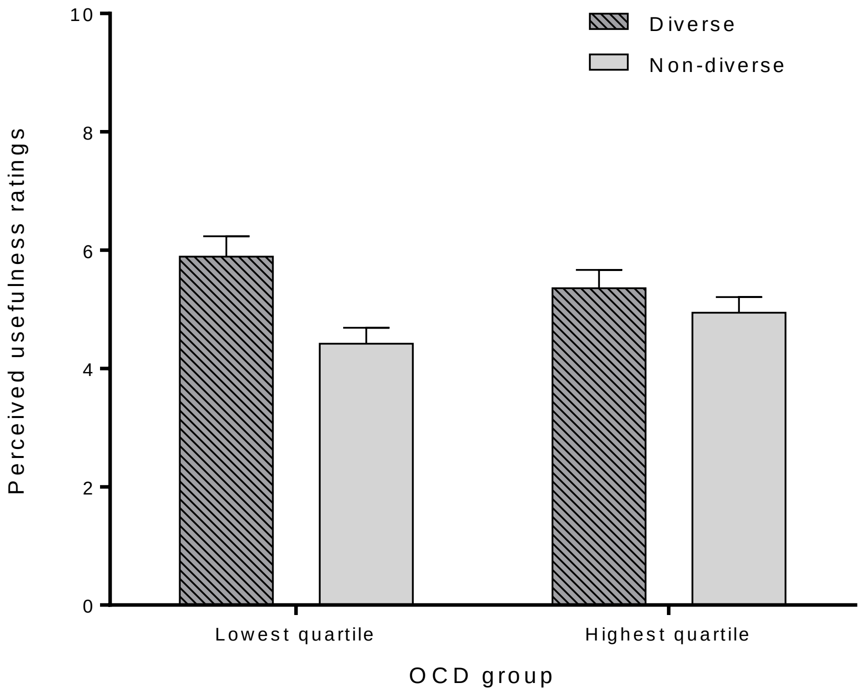

Figure 3.5. Average usefulness ratings (out of 10) for diverse and non-diverse arguments by OCD group on the Experiment 3 forced-choice premise diversity task. Error bars represent standard error of the mean.

Discussion

Experiment 3 utilised a more direct test of the sensitivity of individuals high in OCD symptomatology to the inductive heuristic of premise diversity, relative to those low in OCD symptomatology. This was accomplished by presenting participants with forced-choice decisions between premise pairs that contained diverse or non-diverse evidence in support of a particular conclusion. Under this procedure, individuals highest in OCD symptomatology were less likely to view diverse premise pairs as a stronger basis for inductive generalisation than those lowest in OCD symptomatology. Diverse arguments were also less likely to be rated as more convincing than non-diverse arguments by those highest in OCD symptomatology. These differences in inductive reasoning persisted after factoring out the effects of general ability. Moreover, this appeared to be a global impairment, affecting inductive reasoning about both OCD-relevant and OCD-neutral items.

There are a number of possible mechanisms that may underlie this inductive impairment. Firstly, note that although reliance on diverse information as a basis for inductive generalisations is robust in non-clinical groups (e.g., Kim & Keil, 2003; Kincannon & Spellman, 2003; López, 1995; Osherson et al., 1990), there are some cases in which this inductive heuristic interacts with other factors, such as specific property knowledge. When diverse premises share a highly specific or idiosyncratic property, inductive generalisations based on diverse premises may actually be perceived as weaker than those based on non-diverse premises (Feeney & Heit, 2011). For example, Medin et al. (2003) found that people were less likely to generalise a property shared by *camels* and *desert rats* to other animals, compared to a property shared by *camels* and *rhinos,* even though the first pair was rated as more diverse than the second. In this case, the ecological knowledge that camels and desert rats share the same habitat overrode judgments of inductive strength stemming from premise diversity.

In a similar manner, it is possible that individuals high in OCD symptomatology may infer more specific or idiosyncratic relations between diverse pairs when faced with the forced-choice selection task (cf. O'Connor et al., 2005; Pélissier & O'Connor, 2002), meaning that these diverse arguments are not always seen as a stronger basis for property generalisation. If this were true, one would expect that the diversity effect would be *weaker* in items with OCD-relevant content. However, the weakening of the diversity effect was found equally across both OCD-neutral and OCD-relevant items, suggesting a more general deficit in the processing of premise diversity.

An alternative explanation for this reduced sensitivity to diverse evidence in the highest OCD symptom group is that these individuals perseverate, needing to re-confirm conclusions based on similar premises before exploring the implications of more diverse evidence. Ostensibly, this behaviour appears to relate to the repetitive nature of compulsions, where, for example, individuals with checking compulsions may check locks, windows, and stoves repeatedly. These perseverative tendencies have also been noted on experimental tasks. For example, OCD patients make more perseverative errors on the Wisconsin Card Sorting Task compared to healthy controls (Lucey et al., 1997; Roh et al., 2005; Sanz, Molina, Calcedo, Martin-Loeches, & Rubia, 2001). Additionally, on probabilistic reasoning tasks, OCD patients often require more of similar or repeated information before making judgments (Fear & Healy, 1997; Volans, 1976).

If this explanation accounts, at least in part, for the reduced use of the diversity heuristic, then individuals high in OCD symptoms should repeatedly choose to examine similar types of evidence for a given conclusion, in favour of examining diverse types of evidence. Further, if perseverating, then these individuals should also choose to test more evidence overall, relative to individuals low in OCD symptoms. This possibility was examined in Experiment 4 in a hypothesis testing task, where participants had the opportunity to ‘search’ for additional pieces of evidence that would help to confirm a conclusion.

A third possibility is that the reduced sensitivity to diverse evidence is related to those highest in OCD symptoms actually viewing diverse premises as *more similar* to each other than non-diverse premises. In order for an individual to rely on the diversity of evidence pairs to make inductive judgments, the two premises in a diverse pair must first be recognised as being more distinct from each other as compared to the non-diverse pair. If an individual does not recognise this (i.e., the individual considers the premises in the diverse pair to be more similar than the premises in the non-diverse pair), then responses that may ostensibly seem like a preference for non-diversity may actually represent a diverse response. This resembles Heit et al.'s (2005) explanation for the non-diversity by property reinforcement effect reported by Medin et al. (2003), where traditional diversity effects were reversed when diverse premise categories shared a specific property. Heit et al. (2005) argued that these effects arise because in such arguments, many of the ostensibly 'diverse' premises may be perceived as more similar to one another than the 'non-diverse' premises. This explanation was examined by testing group perceptions of similarity in Experiment 5.

Experiment 4

Experiment 3 established that those lowest in OCD symptoms preferred diverse premise pairs as stronger evidence for a particular conclusion to a greater extent than did those highest in OCD symptoms. One question that follows is whether this preference extends to a more complex and ecologically valid behavioural measure: that of information search when attempting to test a hypothesis. When López (1995) administered such a task to undergraduates, he found that not only was diversity preferred when evaluating categorical arguments, but it was also preferred when the task was to test a hypothesis. López asked participants to select the premise category that would best allow them to test a tentative conclusion, and then to justify their choice. In his experiment, 76% of the participants reported that their choice was based on the dissimilarity of their chosen premise to the given premise. This indicated that people were utilising premise diversity as an explicit reasoning strategy.

Given that in Experiment 3 the participants highest in OCD symptoms exhibited a reduced preference for diverse evidence on an argument evaluation task, this experiment examined whether the same pattern would be found when the task was to search for additional evidence to test a hypothesis. Experiment 4 also afforded the opportunity to examine the possibility of perseverative responding as an explanation for the reduced sensitivity to diversity in those high in levels of OCD symptomatology. If those high in OCD symptoms selected a greater proportion of non-diverse premise pairs in Experiment 3 because they were perseverating on non-diverse evidence, then it is likely that these individuals will repeatedly search for *similar* rather than diverse kinds of evidence in hypothesis testing.

Thus, the task was structured such that participants were given an initial piece of evidence favouring a hypothesis, and were then allowed to test up to six additional pieces of evidence. Two of the search options were highly similar to the given evidence (i.e., non-diverse). The remaining four options showed increasing levels of diversity from the given evidence.

Based on the results of Experiment 3, it was expected that those high in OCD symptoms would be more likely than those low in OCD symptoms to choose non-diverse evidence when testing a hypothesis. Further, under the assumption that the reduced sensitivity to diversity exhibited in Experiment 3 was at least partially explained by individuals high in OCD symptoms being unable to switch from the non-diverse option until they met some threshold of evidence accumulation, it was expected that these individuals would choose to test both non-diverse options first (that is, to test a non-diverse option, followed immediately by the other non-diverse option) more often than those low in OCD symptoms. It was also expected that these individuals would choose to test a greater amount of evidence than would individuals low in OCD symptoms.

Method

Participants

One hundred and seventy-eight undergraduates enrolled in a first-year Psychology course at the University of New South Wales participated in this experiment. Of these, 117 participants also participated in Experiment 7. Consistent with previous experiments, 16 participants (of which 13 also participated in Experiment 7) were excluded because English was not their primary language. All participants received course credit for their participation.

Of the remaining 162 participants, 86 reported high levels of OCD symptoms ($\geq$ 21 on the OCI-R; *M* = 32.02, *SD* = 8.59). Fifty-four of these participants were female, and the average age was 18.86 (*SD* = 1.85). Seventy-six participants reported low levels of OCD symptoms (< 21 on the OCI-R; *M* = 10.51, *SD* = 5.88). Forty-two of these participants were female, and the average age was 19.96 (*SD* = 4.57). The two groups differed on age, with individuals high in OCD symptoms being younger than those in those low in OCD symptoms, $F(1, 160) = 4.21$, $p = .04$, $\eta_p^2 = .03$, but gender distribution did not differ across OCD group, $\chi^2(1, N=162) = 0.96$, $p = .34$.

Materials and Procedure

Participants completed the experiment individually in a quiet room. Of the 104 participants that participated in both the current experiment and Experiment 7, half completed the premise diversity hypothesis testing task, before completing a syllogistic reasoning task (see Chapter 4), whilst the other half completed the tasks in reverse order. The remaining 58 participants only completed the premise diversity hypothesis testing task. All participants then completed the OCI-R and an experimenter administered WASI.

Premise diversity search task. This inductive task was programmed with LiveCode©, and contained eight items. Each item contained a claim and six different category exemplars that could be chosen in the search for further evidence to test the claim. Four of these eight items contained OCD-neutral content (regarding animals or fruit) and the remaining four contained OCD-relevant content (related to contamination and checking concerns). An example of the options presented in an OCD-relevant item is given in Figure 3.6 (see Appendix B for the full item set).

Participants were told to assume that they were scientists who had to test the truth of particular hypotheses by choosing from a set of evidentiary options. First, they were given one piece of evidence that they could assume was true (e.g., “you already know that all dirty tablespoons have traces of the microbe plasmodium vasiporans”). They were instructed to only search as many options as they thought necessary to be ‘reasonably certain’ the hypothesis was true, and to search options in order of informativeness. Specifically, participants were told to “only test as many items as you think is necessary. Please test the items in order of importance (i.e., click the item which you think is most important to test first, the second-most important item second and so on)”.

Text boxes appeared after a selection was made to remind the participants of which options had already been searched and in what order (see Figure 3.6). An example item was given before participants proceeded to the main task. Responses were made via mouse click and unlimited time was given for responses.

In the example presented in Figure 3.6, participants had to evaluate the general claim that ‘all dirty dinnerware has traces of the microbe plasmodium vasiporans’, given that ‘all dirty teaspoons have traces of the microbe plasmodium vasiporans’. They then chose up to a maximum of six evidence options as further tests of the general claim. The testable options were designed to be either dissimilar from the given premise (e.g., bowls and plates), highly similar (e.g., soup spoons and dessert spoons), or of intermediate similarity (e.g., knives and forks, which are still cutlery, but are not types of spoons). There were two options at each level of diversity (i.e., two non-diverse options, two intermediately diverse options, and two diverse options). The order of listing of these options was randomised. Item order was also randomised, with the restriction that one OCD-relevant item and one OCD-neutral item had to appear as the first two items.

After each of the first two items, participants were asked to explain why they had chosen to test those particular evidentiary options. These responses were typed into a text box. Participants were not required to give explanations for the remaining items.

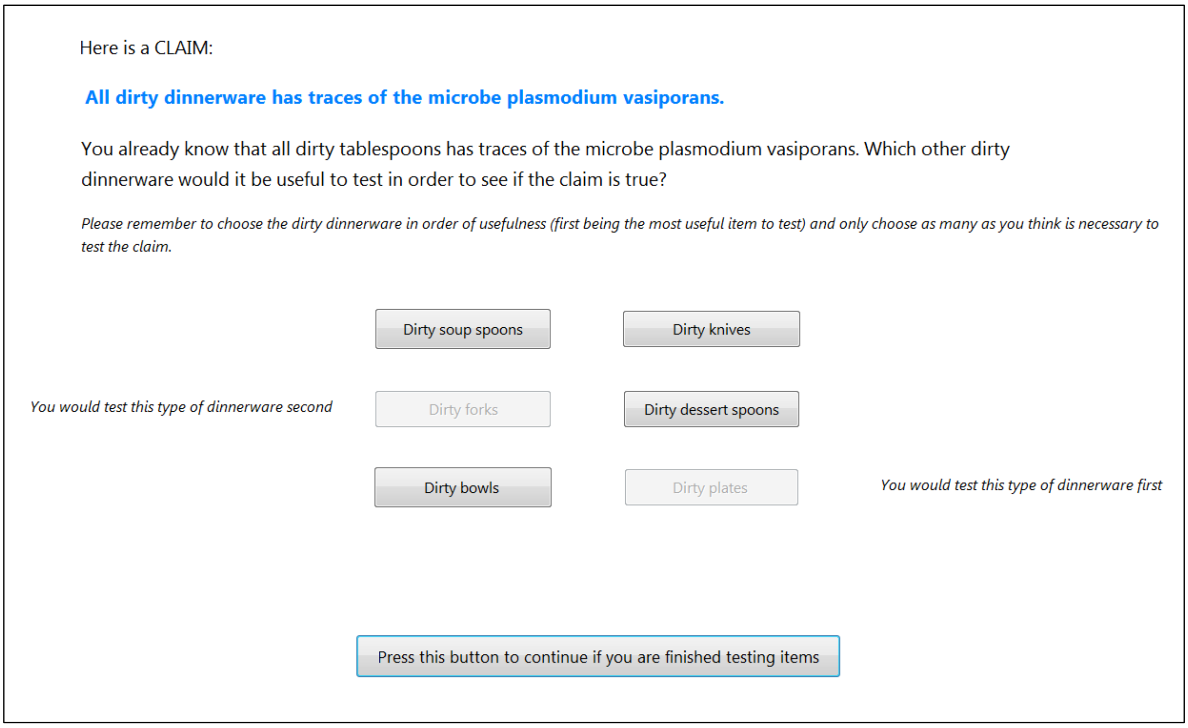

Figure 3.6. Screenshot of an OCD-relevant item used in the Experiment 4 premise diversity hypothesis testing task.

Results

These analyses focused on a number of variables of interest: the mean number of evidence options tested, the proportion of non-diverse or diverse options tested, the proportion of items on which either a non-diverse option or a diverse option was tested first, and the proportion of items on which both non-diverse options were tested as the first two choices. Repeated measures analyses controlling for general ability were used to examine the testing preferences of those high and low in OCD symptoms (see Appendix B for full analyses).

Preliminary Analyses

Preliminary ANOVA analyses indicated that for participants who completed two reasoning tasks, there was no difference of task order (premise diversity hypothesis testing task or syllogistic reasoning task first) on any of the variables of interest (mean number of options tested, proportion of diverse or non-diverse options tested, proportion of items on which a diverse or non-diverse option was tested first, perseveration on non-diverse options; all *F'*s < 2.40). Further, there was no effect of the number of tasks completed (both reasoning tasks or only the premise diversity hypothesis testing task) on any of the variables of interest (all *F'*s < 2.35). Therefore, the remaining analyses were collapsed across both task order and number.

There were differences in general ability between OCD symptom groups. ANOVA analyses indicated that those high in OCD symptoms exhibited significantly lower general ability estimates (M = 104.63, SD = 10.06) compared to those low in OCD symptoms (M = 109.53, SD = 9.55), $F(1, 160) = 10.03$, $p = .002$, $\eta_p^2 = .06$. As in previous studies, full-scale general ability scores were included as a covariate in the analyses.

Premise Diversity Hypothesis Testing Task

Estimated marginal means for the variables of interest can be seen in Table 3.7. For all dependent variables, a 2 (OCD group: high or low) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA was conducted, with repeats on the last factor and controlling for general ability. On average, those low in OCD symptoms searched 3.31 out of a possible 6 options (*SD* = 1.01) and those high in OCD symptoms searched 3.35 (*SD* = 1.02) options. There was no difference between high and low OCD groups on the number of options searched, $F(1, 159) = 0.05$, $p = .83$. The number of options searched also did not differ as a function of OCD-relevance, $F(1, 160) = 0.16$, $p = .69$.

Table 3.7

Estimated Marginal Means for Variables of Interest by OCD Group and OCD-Relevance on the Experiment 4 Premise Diversity Hypothesis Testing Task (Standard Deviations in Parentheses)

Variable of interest	OCD-neutral		OCD-relevant	
	Low OCD	High OCD	Low OCD	High OCD
Overall number tested	3.31 (1.09)	3.33 (1.09)	3.32 (1.08)	3.37 (1.11)
Proportion diverse	0.48 (0.20)	0.44 (0.23)	0.39 (0.12)	0.36 (0.13)
Proportion non-diverse	0.21 (0.21)	0.26 (0.24)	0.25 (0.17)	0.28 (0.18)
Proportion diverse first	0.46 (0.38)	0.44 (0.39)	0.43 (0.30)	0.42 (0.33)
Proportion non-diverse first	0.29 (0.34)	0.34 (0.39)	0.34 (0.32)	0.36 (0.35)
Non-diverse perseveration	0.16 (0.25)	0.21 (0.33)	0.18 (0.26)	0.22 (0.30)

For those low in OCD symptoms, the mean proportion of occasions on which the diverse options were chosen during search was 0.43 (*SD* = 0.15), and for those high in symptoms, this was 0.40 (*SD* = 0.17); there was no reliable difference between high and low OCD groups in the overall proportion of diverse options chosen, $F(1, 159) =$

1.81, $p = .18$. The proportion of occasions on which a non-diverse option was chosen during the search was 0.23 ($SD = 0.17$) for those low in symptoms and 0.27 ($SD = 0.20$) for those high in symptoms; this proportion did not differ between groups, $F(1, 159) = 1.58$, $p = .21$[7].

However, OCD-relevance did affect search choices; as shown in Table 3.7, participants were more likely to choose diverse options when item content was OCD-neutral than when it was OCD-relevant, $F(1, 160) = 50.94$, $p < .001$, $\eta_p^2 = .24$, but were more likely to choose non-diverse options when item content was OCD-relevant than when it was OCD-neutral, $F(1, 160) = 9.39$, $p = .003$, $\eta_p^2 = .06$.

Both groups high and low in OCD symptoms chose diverse evidence as their first search option more often than would be expected by chance (low-OCD, $t(85) = 2.11$, $p = .04$; high-OCD, $t(75) = 3.80$, $p < .001$). Rates of initial search of diverse options did not differ between OCD symptom groups, $F(1, 159) = 0.08$, $p = .77$. There were also no OCD group differences in the rate of initial search of non-diverse options, $F(1, 159) = 0.42$, $p = .52$. Overall, the proportion of non-diverse options tested first did not differ from the level expected by chance for either OCD status group (low-OCD, $t(85) = 1.09$, $p = .28$; high-OCD, $t(75) = -0.99$, $p = .33$).

One hypothesised explanation for the reduced sensitivity to premise diversity observed in Experiment 3 was that individuals high in OCD symptoms needed to re-confirm conclusions by examining similar premises before they explored the implications of more diverse evidence (i.e., they perseverate on non-diverse information). There was, however, no difference between OCD groups on the proportion of items on which non-diverse perseveration was observed, $F(1, 159) = 1.12$,

[7] Note that when general ability was not controlled, those high in OCD symptoms searched fewer diverse options (low-OCD, $M = 0.44$, $SD = 0.15$; high-OCD, $M = 0.39$, $SD = 0.17$), $F(1, 160) = 4.33$, $p = .04$, $\eta_p^2 = .03$, and a greater number of non-diverse options (low-OCD, $M = 0.22$, $SD = 0.17$; high-OCD, $M = 0.28$, $SD = 0.19$), $F(1, 160) = 4.08$, $p = .05$, $\eta_p^2 = .03$, relative to those low in OCD symptoms.

p = .29. That is, those high and low in OCD symptoms did not differ in the proportion of items on which non-diverse options were tested first and second in the search sequence. No other main effects or interactions reached significance, and the pattern of results did not change when comparisons were restricted to those highest and lowest in OCD symptoms (see Appendix B for these analyses).

Explanations of search choices. Self-reported explanations regarding why participants had chosen particular evidentiary options to test on the first OCD-relevant and the first OCD-neutral item were binary coded by the author, based on the approach developed by López (1995). Explanations were coded as 'diverse' if they mentioned the dissimilarity of the option to the given premise information as a reason for selection, either directly (e.g., "I chose fruit/s X because it is different from an orange") or indirectly (e.g., "I chose fruit/s X because it is not a citrus fruit"). Explanations that did not mention the dissimilarity/diversity of the chosen option were coded as 'other'. Explanations from a randomly selected 25% of the sample (including both individuals high and low in OCD symptomatology) were re-coded separately by an independent judge. Inter-rater agreement was over 95% for both OCD-neutral and OCD-relevant items.

For OCD-neutral items, 63% of individuals low in OCD symptoms and 50% of individuals high in OCD symptoms either directly or indirectly mentioned item diversity in their explanation of search choices. This group difference in the distribution of responses approached significance, $\chi^2(1, N=162) = 3.35$, $p = .08$. For OCD-relevant items, 43% of those low in OCD symptoms and 42% of those high in OCD symptoms directly or indirectly mentioned dissimilarity as a reason for their choices. This distribution did not differ across groups, $\chi^2(1, N=162) = 0.04$, $p = .87$.

Discussion

This experiment aimed to determine whether the reduced preference for diverse premises exhibited by those highest in OCD symptoms found in Experiment 3 would extend to a hypothesis testing task. Both high and low OCD symptom groups were more likely than chance to test the diverse option first. This replicates López's (1995) finding that adults prefer diverse evidence when selecting which evidence options represent the strongest tests for a particular claim. Contrary to expectations, however, no differences were found between high and low OCD groups in the overall proportion of occasions on which the diverse option was tested first, nor were any group differences noted on the proportion of items on which a non-diverse option was tested first. For OCD-neutral items, there was a trend toward those low in OCD symptoms mentioning diversity as a justification for the options searched more often than those high in OCD symptoms.

There were also no OCD group differences in the overall proportion of diverse or non-diverse options searched. Further, no OCD group differences were noted on the proportion of perseverative responses (choosing both non-diverse options as the first and second search options), or in the overall amount of evidence tested. The unexpected failure to find group differences on these measures seems to indicate that those high in OCD symptoms may be as able to utilise diversity as those low in symptoms when searching for evidence to support hypotheses. Moreover, these results suggest that the OCD group differences in sensitivity to diversity found in Experiment 3 were not due to a perseverative focus on similar types of evidence.

Why were OCD group differences in sensitivity to diversity found on the forced-choice task and not the hypothesis testing task? First, note the differences in the respective task requirements. Responding in the Experiment 3 forced-choice task involved argument *evaluation,* as participants judged whether premise pairs were strong or weak evidence for a particular conclusion. In this task, there was no need to search for further evidence. In contrast, the current task focused on argument *testing*, which required searching the available premises for further evidence to test a given conclusion. Unlike the binary choice task where participants were told that all premises could be assumed to be true, in the search task this remained ambiguous. Hence, some participants may have searched particular options because they believed it was a way to confirm the claim, whilst others may have thought it was the best way of disconfirming the claim. This ambiguity makes it difficult to know exactly why people chose diverse search options. It may be that they believed a property shared by two or more diverse cases was a better basis for confirming a hypothesis. However, if one does not always subscribe to the diversity heuristic (as suggested would be the case for high OCD symptom individuals) *and* believes that some premises are false, then choosing the diverse option is a good way of disconfirming the conclusion (i.e., choosing the diverse option is more likely to demonstrate that the conclusion is not true for all cases).

A further difficulty experienced with the hypothesis testing task is that with so many testing options (order of options tested, number of options, diversity of options), the dataset was necessarily complex and noisy. For example, as the proportion of items on which a non-diverse option was tested first was relatively low, it is difficult to answer the question of whether those high in OCD symptoms were perseverating on the non-diverse option. A future iteration of the experiment could improve on this by simplifying the number of available search options. This may be accomplished by only including the most extreme non-diverse and diverse options (and removing the options in between).

Experiment 5

One other explanation that may account for the reduced use of the diversity heuristic in Experiment 3 are group differences in perception of the similarity of the premises used. Recall that the diversity of premise pairs necessarily depends on how similar these premises are to each other. In the forced-choice premise diversity task, each item was chosen based on pre-test ratings of the relative similarity of the premises. The OCD status of those giving the similarity ratings was not assessed. However, based on the distributions of OCD symptoms found in Experiments 1 to 4, it seems safe to assume that those generating the similarity ratings were likely to be low in OCD symptoms.

If those high in OCD symptoms perceive differences between premises where others perceive similarities, or vice versa, this may account (at least partially) for their reduced reliance on the diversity heuristic. To test this possibility, an online experiment on Amazon’s Mechanical Turk (MTurk) system was conducted, to explicitly compare perceptions of similarity in those high and low in OCD symptomatology.

Whilst MTurk is a relatively new system for collecting experimental data, research comparing MTurk participant performance to more traditional samples supports the notion that the data produced by MTurk samples are reliable. For example, individual difference self-report data have been noted to be psychometrically valid (Buhrmester, Kwang, & Gosling, 2011; Shapiro, Chandler, & Mueller, 2013). In addition, well-established cognitive biases (e.g., framing effects), logical facilities (e.g., conjunction fallacies), and patterns on cognitive reaction time tasks (e.g., Stroop) have been replicated in MTurk samples (e.g., Crump, McDonnell, & Gureckis, 2013; Goodman, Cryder, & Cheema, 2013; Horton, Rand, & Zeckhauser, 2011; Paolacci, Chandler, & Ipeirotis, 2010). MTurk participants are highly intrinsically motivated (e.g., participating because they consider tasks to be recreational), and the incentive structure of MTurk encourages diligence, as experimenters only approve submissions that satisfy certain standards and can screen participants on the basis of past approval rates (Paolacci & Chandler, 2014).

As the current experiment was delivered on an online platform, the experimenter-administered WASI could not be used to estimate general ability. Thus, a complex vocabulary task was administered to participants to serve as a proxy measure of general ability. This approach is justified because scores on vocabulary subtests load highly onto general ability (Brody, 1992; A. S. Kaufman, 1990; Wechsler, 1997, 1999). Specifically, a word identification task was employed, which required participants to discriminate between complex real words and non-words (Stanovich & West, 1998b; Zimmerman, Broder, Shaughnessy, & Underwood, 1977). This task reliably assesses differences in individual vocabulary knowledge (Anderson & Freebody, 1983; Cooksey & Freebody, 1987) and correlates highly with SAT scores (Stanovich & West, 1998b; Zimmerman et al., 1977).

Method

Participants

Two hundred and seventy participants participated in this experiment on the Amazon MTurk system. Participants were from the United States, and were screened so that they met an approval rate of at least 90% on their past submissions, as participants with high approval ratings produce higher quality data than low-reputation participants (Peer, Vosgerau, & Acquisiti, 2013).

All participants received $1 USD in return for their participation. Of the 270 participants that completed the task, one was excluded as English was not his primary language. Participants were split into OCD-status groups according to their scores on the OCI-R (low OCD symptom group, <21 on the OCI-R; high OCD symptom group, ≥ 21 on the OCI-R). One hundred and ninety-nine participants were in the low OCD symptom group (M = 9.11, SD = 5.49); of these, 101 were female, and the mean age was 37.13 years (SD = 12.74). There were 70 participants in the high OCD symptom group (M = 30.00, SD = 10.63); 35 participants in this group were female, and the mean age was 32.84 years (SD = 11.89). The low OCD symptom group was older than the high OCD symptom group, $F(1, 267) = 6.07$, $p = .01$, $\eta_p^2 = .02$, but there was no difference in gender distribution across the OCD-status groups, $\chi^2(1, N=269) = 0.01$, $p = 1.00$.

Materials and Procedure

The experiment was designed on Qualtrics (www.qualtrics.com), an online survey platform. Participants completed the similarity rating task before completing the OCI-R. To finish the experiment, participants completed the vocabulary task.

Similarity ratings. Participants were presented with the 30 premise pairs (divided equally into OCD-neutral and OCD-relevant content) utilised in Experiment 3. They were given a premise (e.g., ‘all silver coins are contaminated by the germ *hemonasella coli*’), and had to choose which of two probe premises were more similar to the given premise (e.g., ‘all gold coins are contaminated by the germ *hemonasella coli*’ or ‘all dollar bills are contaminated by the germ *hemonasella coli*’). The given premise was always the premise that was common to both diverse and non-diverse items in the forced-choice diversity task used in Experiment 3. Conclusions were not presented.

Following this choice, participants rated the similarity of each of the two probes to the given claim on a Likert-rating scale from 0 (not at all similar) to 10 (extremely similar). Item presentation order was randomised, as was the order in which the probes were presented within each item. Participants made ratings via mouse click on an on-screen scale for the 30 target items. Three attention check items were included: these were the same items as presented in Experiment 2, where the conclusion category was completely unrelated to the premises (e.g., premises relating to mammals and a conclusion relating to alcohol).

Word identification task. Participants saw a list of 100 complex words that included 74 real words and 26 non-words taken from Zimmerman, Broder, Shaughnessy, and Underwood (1977). Participants were asked to indicate whether a word was a real word (e.g., insculp) or whether it was a non-word (e.g., ineffity). In line with the procedure used by Stanovich and West (1998b), a word identification accuracy score was calculated by subtracting the proportion of foil responses (responses that indicated that a non-word was a word) from the proportion of hits (responses that indicated a real word was a word).

Results

Eight participants were excluded from the analyses because they answered at least one of the three attention check questions incorrectly. Note that quartile group analyses were not carried out due to the relatively low number of individuals with OCI-R scores of 21 or over (see Appendix B for full OCD group analyses).

Preliminary Analyses

Those high in OCD symptoms (M = 0.27, SD = 0.14) showed less accurate discrimination between non-words and real words on the word identification task, relative to those in the low in OCD symptoms (M = 0.34, SD = 0.14), $F(1, 268)$ = 12.09, p = .001, η_p^2 = .04. This is line with Experiments 3 and 4, which found that individuals low in OCD symptomatology had higher general ability scores than individuals high in OCD symptomatology.

For neutral items, there was a weak but marginally reliable correlation between word identification accuracy and the likelihood that the pattern of similarity judgments differed in the current experiment from the Experiment 2 pre-test (i.e., the proportion of times that pairs previously endorsed as dissimilar were endorsed as *more* similar in this experiment), $r(259) = -.12$, $p = .05$; this was not the case for OCD-neutral items, $r(259) = -.04$, $p = .57$. Hence, in subsequent analyses of similarity ratings, word identification accuracy was used as a proxy for general ability and was included as a covariate.

Similarity Ratings Task

Proportion of previously nominated 'non-diverse' items endorsed as similar ('matching' responses). As can be seen in Table 3.8, both high and low OCD symptom groups were more likely to endorse the 'non-diverse' premises from Experiment 3 as being the more similar premise pair than would be expected by chance; this was true of both OCD-neutral and OCD-relevant content (all *p*'*s* < .001).

Table 3.8

Estimated Marginal Mean Proportion of 'Matching' Responses (Nominally 'Non-Diverse' Items Rated as Similar) by OCD Group and OCD-Relevance on the Experiment 5 Similarity Rating Task (Standard Deviations in Parentheses) and Details of Significance Tests Compared to Chance Responding (0.50)

	Low OCD group				High OCD group			
OCD-relevance	Proportion 'matching'	*t*	*df*	*p*	Proportion 'matching'	*t*	*df*	*p*
Neutral	0.99 (0.04)	198.42	196	<.001**	0.96 (0.14)	27.23	63	<.001**
Relevant	0.92 (0.08)	79.57	196	<.001**	0.86 (0.14)	20.52	63	<.001**

Note. *p < .05, **p < .01.

A 2 (OCD group: high or low) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA with repeats on the last factor and controlling for proxy general ability estimates was conducted on the proportion of previously nominated 'non-diverse' items that were endorsed as being the most similar option. Henceforth, these responses are referred to as 'matching' responses (i.e., responses that 'match' the assignment of premises to either non-diverse or diverse pairings in Experiment 3).

Overall, those low in OCD symptoms (M = 0.96, SD = 0.05) were more likely to give matching responses than those high in OCD symptoms (M = 0.91, SD = 0.13), F (1, 258) = 18.99, p < .001, η_p^2 = .07. That is, individuals high in OCD symptoms were more likely to indicate that the previously nominated 'non-diverse' items were more *dissimilar* to a given premise relative to those low in symptoms. An increased number of matching responses was given to OCD-neutral items over OCD-relevant items, F (1, 259) = 217.51, p <.001, η_p^2 = .46. Notably, this effect of OCD-relevance interacted with OCD group, F(1, 259) = 11.17, p = .001, η_p^2 = .04. Bonferroni-adjusted simple effect analyses with an alpha level of 0.125 indicated that for both OCD-neutral, F(1, 258) = 8.27, p = .004, η_p^2 = .03, and OCD-relevant content, F(1, 258) = 22.08, p < .001, η_p^2 = .08, those low in OCD symptoms were more likely to offer matching responses than those high in OCD symptoms, but this effect was larger for the OCD-relevant items.

Similarity ratings. A 2 (OCD group: high or low) × (2) ('diversity' status: previously nominated 'diverse' or 'non-diverse') × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA with repeats on the last two factors and controlling for proxy general ability estimates was also conducted on similarity ratings given to the premises. Estimated marginal means for similarity ratings can be seen in Table 3.9.

Table 3.9

Estimated Marginal Mean Similarity Ratings (out of 10) Given to Previously Nominated 'Diverse' and 'Non-Diverse' Premises by OCD Group on the Experiment 5 Similarity Rating Task (Standard Deviations in Parentheses)

OCD-relevance	Low OCD group		High OCD group	
	'Diverse'	'Non-diverse'	'Diverse'	'Non-diverse'
Neutral	3.64 (1.99)	7.56 (1.15)	3.92 (1.99)	7.40 (1.23)
Relevant	5.13 (1.57)	7.84 (0.93)	5.21 (1.55)	7.71 (0.98)

As can be seen in Table 3.9, previously nominated 'non-diverse' premises were rated as more similar to the given premise than 'diverse' premises, $F(1, 259) = 671.89$, $p < .001$, $\eta_p^2 = .72$. Overall, participants gave higher similarity ratings to premises that were OCD-relevant than OCD-neutral, $F(1, 259) = 283.79$, $p < .001$, $\eta_p^2 = .52$. However, there was also an interaction between the previously nominated diversity-status and OCD-relevance of the premises. Table 3.9 shows that there was a much larger discrepancy between similarity ratings for diverse and non-diverse premises for OCD-neutral items than for OCD-relevant items, $F(1, 259) = 227.43$, $p < .001$, $\eta_p^2 = .47$. There was no main effect of OCD group on similarity ratings, $F(1, 258) = 0.27$, $p = .60$, nor did OCD group interact with either the diversity-status of the items, $F(1, 259) = 0.23$, $p = .63$, or the OCD-relevance of item content, $F(1, 259) = 1.01$, $p = .32$.

Discussion

This experiment aimed to determine whether individuals high in OCD symptoms perceived similarities between argument premises in the same way as individuals low in OCD symptoms. The forced-choice similarity task suggested that there were reliable group differences in perceptions of similarity. Those high in OCD symptoms were less likely to perceive relative similarity in a way that matched the previous classification of similar/non-diverse and dissimilar/diverse items in

Experiment 2. This was true for both OCD-neutral and OCD-relevant content, although group differences in perception of similarity were larger for OCD-relevant content.

It is unclear why these group differences in the perception of similarity were not also apparent across similarity ratings. Note, however, that the overall rate of mismatching responses was low. On average, those high in OCD symptoms only endorsed mismatching (i.e., endorsing the 'diverse' option as more similar) responses less than 5% of the time on OCD-neutral items, and less than 15% of the time on OCD-relevant items. Presumably then, the reversed similarity ratings (i.e., higher similarity ratings for nominally 'diverse' items) given to these items were diluted by the similarity ratings given to the vast majority of items on which the perception of similarity did not differ between the two OCD groups.

How can group differences in the perception of similarity be explained? First, note that similarity is highly flexible and dependent on knowledge and purpose (Goldstone & Son, 2005). That is, the similarity that one perceives between two items depends strongly on the knowledge that one possesses about those items. For example, when asked to sort mammals that "go together by nature", rural American undergraduates sorted on the basis of size and ferocity, whereas adults from an Itzaj-Mayan culture gave greater weight to ecological considerations, such as habitat and diet, and clustered mammals into smaller groups (López et al., 1997). Similarly, commercial fishermen employed ecological, commercial, and behavioural knowledge to sort marine creatures into categories, whereas university undergraduates did not (Shafto & Coley, 2003). Both these studies suggest that perceived similarity is highly dependent on background knowledge.

By extension, it is plausible that individuals high in OCD symptoms rely on different considerations to individuals low in OCD symptoms when making their judgments of similarity. Whilst it is difficult to account for why those high in OCD symptoms also perceived greater similarity between the given and ‘diverse’ premises for OCD-neutral content, it is easier to account for group differences in the perceived similarity of OCD-relevant content. Presumably those high in OCD symptoms have greater experience with (or are at least more concerned by) the content of OCD-relevant items, which may have resulted in the formation of a somewhat different conceptual structure of OCD-relevant knowledge. This corresponds with the differences in perceived similarity between OCD groups being larger for OCD-relevant items compared to OCD-neutral items.

An implication of this finding is that the apparent reduced preference for diverse evidence noted in the forced-choice diversity task (Experiment 3) may not be a reduced preference for diversity at all. Sometimes non-diverse premises are seen as a stronger basis for inductive generalisation relative to diverse premises because these non-diverse premises are actually perceived as *less* similar than diverse premises; this still constitutes a diversity effect (Heit et al., 2005). Thus, individuals high in OCD symptoms may still be basing their inductive judgments on the perceived dissimilarity between premises (i.e., reasoning according to diversity), but these individuals may perceive more similarities between ‘diverse’ premise pairs (or alternatively dissimilarities between ‘non-diverse’ pairs), which may have the appearance of a reduced sensitivity to premise diversity.

However, re-visiting the results of Experiment 3 suggests that this explanation is unlikely to entirely account for group differences in sensitivity to diversity. Firstly, the reduced sensitivity to premise diversity was a global effect equal across *both* OCD-neutral and OCD-relevant items. In contrast, although those high in OCD symptoms gave fewer matching responses across both OCD-neutral and OCD-relevant items, this effect was much larger for OCD-relevant items.

To further explore this issue, an analysis comparing the sensitivity of quartile groups to premise diversity on OCD-relevant items in Experiment 3 was re-run. Items on which individuals high in OCD symptoms gave mismatching similarity responses at least 10% more frequently than those low in OCD symptoms were dropped from this analysis. This left a subset of 11 (out of 15) items where the forced-choice similarity judgments of those high in OCD symptoms were in general agreement with the ratings in the Experiment 3 pre-test. In this re-analysis, those highest in OCD symptoms (M = 0.51, SD = 0.25) *still* exhibited a reduced preference for inductive arguments based on diverse premises, relative to those with the lowest OCD symptoms (M = 0.62, SD = 0.19) whilst controlling for general ability, $F(1, 59) = 5.09$, $p = .03$, $\eta_p^2 = .08$. Thus, it appears that differences in perception of similarity are unlikely to entirely account for the reduced sensitivity to diversity exhibited by those highest in OCD symptoms.

Summary

Collectively, Experiments 2 to 5 have suggested that individuals high in OCD symptoms are: 1) less sensitive to argument validity as defined by class-inclusion relationships, and 2) less sensitive to the implications of argument diversity, relative to individuals low in OCD symptoms. That is, deficits in both deductive and inductive reasoning were noted for those high in OCD symptoms. Importantly, neither of these impairments were accounted for by group differences in general ability.

Although the source of the reduced sensitivity to premise diversity has not been entirely isolated, two explanations for this effect are unlikely. It is unlikely that those high in OCD symptoms are: 1) perseverating on similar (non-diverse) types of evidence, and 2) employing other forms of causal or ecological knowledge to make their inductive generalisations. However, a reduced sensitivity to diversity may be at least partially (but not completely) explained by differences in the way that high and low OCD symptom groups perceive similarity. These issues and their theoretical and clinical implications are examined in more detail in the General Discussion (Chapter 6).

CHAPTER 4 - Sensitivity to Background Knowledge and Argument Validity in Obsessive-Compulsive Disorder

The group reasoning differences between individuals high and low in OCD symptoms found so far challenge the "spared deduction, impaired induction" account of OCD. Examination of the inductive portion of this account has focused primarily on the inductive sampling heuristics of argument length (premise monotonicity) and premise diversity. Although individuals high in OCD symptomatology did exhibit a reduced preference for diverse evidence relative to those low in OCD symptomatology (Experiments 2 and 3), these individuals appeared unimpaired in their use of sample size to make judgments of inductive strength (Experiment 1). Thus, the picture so far of the ability of individuals high in OCD symptoms to employ inductive sampling heuristics has been mixed.

Beyond inductive heuristics, inductive reasoning can also be influenced by existing background knowledge (cf. Hayes et al., 2010 for a review). It is an open question whether induction based on background knowledge is impaired for those high in OCD symptoms relative to those low in OCD symptoms. This was addressed in Experiment 6. Moreover, although deductive reasoning depends primarily on the application of logical rules, background knowledge can also influence deductive judgments, especially through alteration of the believability of argument conclusions (Evans, 1989; Evans et al., 1983; Newstead & Evans, 1993). This influence of background knowledge on the deductive performance of individuals high and low in OCD symptoms was examined in Experiment 7.

In terms of deductive reasoning, individuals high in OCD symptomatology have been shown to be poorer at distinguishing between valid and invalid arguments, relative to those low in symptomatology (Experiments 1 and 2). Importantly, these differences have been found even when controlling for general ability, and across both OCD-neutral and OCD-relevant content. So far in this series of experiments, however, deductive performance has only been examined under a relatively elementary manipulation of logical structure: class-inclusion, where valid arguments include a premise about an overarching category (e.g., mammals, flowers) possessing a particular property and invalid arguments do not.

It remains to be seen whether these OCD group differences in deductive reasoning persist in arguments with a different logical structure. There is good reason to believe that this is so: when reasoning about a syllogistic reasoning task with conclusions that varied in validity, believability, and OCD-relevance, OCD patients were poorer at distinguishing between valid and invalid syllogisms on OCD-neutral items, relative to general ability and age-matched non-clinical controls (Simpson et al., 2007). An example of an invalid/believable OCD-neutral syllogism was ‘No well-educated people are fashionable. Some judges are fashionable. Does it follow that some well-educated people are not judges?’ In contrast, OCD patients were as able as non-clinical controls to discriminate between valid and invalid items when reasoning with OCD-relevant items (e.g., an invalid/believable OCD-relevant syllogism was ‘No life-threatening cancers are detectable. Some cancers are detectable. Does it follow that some life-threatening diseases are not cancers?’).

Deductive reasoning according to different logical structures was examined in two separate experiments. Experiment 6 adapted the induction-deduction task devised by Rips (2001) to analyse the influence of background knowledge on both induction and deduction. Experiment 7 employed a deductive reasoning task similar to the task used by Simpson et al. (2007), in order to examine the influence of background knowledge on deductive reasoning.

Experiment 6

In line with the paradigm developed by Rips (2001), participants were presented with a common item set and asked to judge the items according to either deductive validity or inductive plausibility. This facilitated the direct comparison of deductive and inductive reasoning, and limited the confounds that usually interfere with attempts to compare the two reasoning processes using dissimilar tasks (e.g., task difficulty, item content).

The item set that Rips (2001) presented to participants was fully crossed according to validity (i.e., whether conclusions were *necessarily* true given only the premises) and causal consistency[8] (i.e., whether item content was consistent with background cause and effect knowledge), resulting in four different types of items: valid/consistent, valid/inconsistent, invalid/consistent and invalid/inconsistent (see Table 4.1). Rips found that under deduction conditions, binary judgments of argument strength were primarily influenced by deductive validity, regardless of causal consistency. In contrast, those given induction instructions were highly sensitive to causal consistency. In particular, invalid/consistent arguments (e.g., top right cell of Table 4.1) were judged to be of greater argument strength than valid/inconsistent arguments (e.g., bottom left cell of Table 4.1). Rips argued that this indicated that people were using qualitatively different criteria for evaluating arguments under induction and deduction instructions (see Singmann & Klauer, 2011 for similar results).

[8] Note that this definition of 'causal consistency' is not stringent, and may include background knowledge that is not strictly causal in nature.

Table 4.1

Examples of the Four Argument Types Used in the Experiment 6 Causal Consistency Induction-Deduction Task

	Argument validity	
Causal status	Valid	Invalid
Causally consistent	If car #10 runs into a brick wall, it will stop Car #10 runs into a brick wall ______________________ *Car #10 stops*	Car #10 runs into a brick wall ______________________ *Car #10 stops*
Causally inconsistent	If car #10 runs into a brick wall, it will speed up Car #10 runs into a brick wall ______________________ *Car #10 speeds up*	Car #10 runs into a brick wall ______________________ *Car #10 speeds up*

Note. The premises are given in normal font above the line and are assumed to be true. Conclusions are given in italics below the line. An argument is ‘strong’ if the conclusion is plausible given the premises, and is ‘valid’ if necessarily true.

It was expected that individuals low in OCD symptoms would exhibit a similar response pattern to the participants in Rips (2001). That is, under deduction instructions, these individuals would be primarily influenced by argument validity, regardless of the causal consistency of item content. In contrast, under induction instructions, those low in OCD symptoms were expected to be greatly influenced by the consistency of item content with background knowledge.

Individuals high in OCD symptomatology were expected to show a different pattern. These individuals were expected to be impaired in their deductive judgments (e.g., Experiments 1 and 2; Simpson et al., 2007). Specifically, to the extent that those high in OCD symptoms are impaired in their deductive judgments across arguments with different logical forms, these individuals were expected to be less able to discriminate between valid and invalid arguments, relative to individuals low in OCD symptoms. Further, relative to those low in OCD symptoms, those high in OCD symptoms were not expected to be as sensitive to the causal consistency of arguments under induction instructions, as they have previously been shown to be impaired when making inductive judgments (e.g., Experiments 2 and 3; Aardema et al., 2009; Pélissier & O'Connor, 2002). Based on the findings of Experiments 2 and 3, it was expected that these impairments would be robust over and above any group differences in general ability. Note that the OCD-relevance of item content was not manipulated in the current experiment, as this was one of the earliest experiments run in the course of this thesis.

Method

Participants

One hundred and eleven undergraduates enrolled in a first-year Psychology course at the University of New South Wales participated. They also participated in Experiment 3. All were recruited via an online experimental system and received course credit for their participation. Consistent with the other experiments in this thesis, 11 participants were excluded from the experiment because English was not their primary language.

Participants were randomly assigned to either the deduction or induction condition. There were 50 participants in each reasoning condition. These participants were separated into groups based on their scores on the OCI-R (Foa et al., 2002). Forty-four participants scored equal to or greater than 21 on the OCI-R and were defined as high OCD symptom participants (*M* = 29.95, *SD* = 7.19); of these, 30 participants were female with a mean age of 18.75 (*SD* = 1.64). Fifty-six individuals scored below 21 on the OCI-R and were defined as low OCD symptom individuals (*M* = 10.86, *SD* = 5.42); 30 of these participants were female, with a mean age of 20.46 years (*SD* = 6.67). The two groups did not differ by age, $F(1, 98) = 2.77$, $p = .10$, nor in the distribution of gender across groups, $\chi^2(1, N=100) = 0.14$, $p = .16$.

Materials and Procedure

Participants were tested individually in a quiet room. Half of the participants began the experiment with the causal consistency induction-deduction task, before completing the forced-choice premise diversity task (see Chapter 3, Experiment 3). The remainder completed the tasks in reverse order. All participants then completed a computerised version of the OCI-R. Finally, all participants completed the two-subtest version of the WASI, administered by the author.

Causal consistency induction-deduction task. The induction-deduction task was programmed using LiveCode©. A set of 16 arguments was created: half were based on the modus ponens structure (If *p* then *q*, *p* therefore *q*) and half were based on the conjunctive syllogism structure (not (*p* and *q*), *p* therefore not *q*). Arguments varied in validity, such that half were logically valid (e.g., in the left-hand column of Table 4.1) and half were logically invalid (e.g., in the right-hand column of Table 4.1). They also varied in causal consistency: half contained content consistent with background knowledge (e.g., in the top row of Table 4.1), while the remainder were inconsistent with such knowledge (e.g., in the bottom row of Table 4.1). The majority of these items were adapted from Rips (2001) and were presented in a randomised order (see Appendix C for the full stimuli set). All items were deemed OCD-neutral in content.

To begin, those in the deduction condition saw definitions of 'valid' and 'invalid' arguments. In line with Experiments 1 and 2, valid arguments were defined as those that led to a conclusion that was *necessarily true*, whereas invalid arguments were those that led to a conclusion that was *not necessarily true*, given that the premises were true. Those in the induction condition saw definitions of 'strong' and 'weak' arguments.

Strong arguments were defined as those for which the conclusion was *plausible* given the premises, whereas weak arguments were those where the conclusion was *implausible*.

Participants made binary judgments regarding whether each of the arguments presented were ‘valid’/ ‘invalid’ or ‘strong’/ ‘weak’. These decisions were made via mouse click, and were untimed, with arguments remaining on the screen until participants responded. After each judgment, participants indicated how confident they were in their decision on a Likert scale of 1 (not confident at all) to 5 (very confident).

Results

These analyses examined the proportion of positive responses (i.e., the proportion of items that were evaluated as ‘valid’ under deduction instructions or ‘strong’ under induction instructions in binary judgments) in each reasoning condition and OCD symptom group. Following this, confidence ratings were examined (see Appendix C for full analyses).

Preliminary Analyses

ANOVA analyses compared binary ‘valid’/ ‘strong’ judgments in the counterbalanced task order conditions (induction-deduction task first or premise-diversity task first). There was no effect of task order, $F(1, 91) = 0.34$, $p = .56$. Therefore, all subsequent analyses were collapsed across this factor.

ANOVA analyses also compared the proportion of binary ‘valid’/ ‘strong’ judgments to different argument structures (modus ponens or conjunctive syllogisms). Overall, participants were more likely to endorse modus ponens arguments ($M = 0.61$, $SD = 0.18$) than conjunctive syllogisms ($M = 0.51$, $SD = 0.19$), $F\ (1, 96) = 16.33$, $p < .001$, $\eta_p^2 = .15$. Validity interacted with argument structure, $F(1, 96) = 23.60$, $p < .001$, $\eta_p^2 = .20$; sensitivity to validity (as defined by proportion of positive responses to valid items – proportion of positive responses to invalid items) was larger for modus ponens arguments (*M* = 0.51, *SD* = 0.35) than for conjunctive syllogisms (*M* = 0.25, *SD* = 0.39). Argument structure did not interact with causal consistency, $F(1, 96) = 2.81$, $p = .10$, OCD group, $F(1, 96) = 0.01$, $p = .91$, or reasoning condition, $F(1, 96) = 1.31$, $p = .26$. Collectively, this seems to indicate that conjunctive syllogisms were more difficult than modus ponens arguments. However, as argument structure did not interact with any other factor, subsequent analyses are collapsed across this factor.

A one-way ANOVA confirmed that high and low OCD symptom groups did not differ in general ability estimates, as measured by the WASI, $F(1, 98) = 3.36$, $p = .07$. However, those scoring in the highest quartile of OCD symptoms had lower general ability scores (*M* = 105.37, *SD* = 1.91) than those scoring in the lowest quartile of OCD symptoms (*M* = 112.42, *SD* = 2.03), $F(1, 49) = 6.41$, $p = .02$, $\eta_p^2 = .12$.

Causal Consistency Induction-Deduction Task

Proportion of positive responses. These analyses examined the proportion of positive (‘strong’/ ‘valid’) responses to all four argument types (valid/consistent, valid/inconsistent, invalid/consistent and invalid/inconsistent). Estimated marginal means for these items can be seen in Figure 4.1. The proportion of positive responses was analysed in a 2 (level of OCD-symptoms: high or low) × 2 (reasoning condition: deduction or induction) × (2) (item-validity: valid or invalid) × (2) (item-causal consistency: causally consistent or causally inconsistent) ANCOVA with repeated measures on the last two factors and controlling for general ability.

The analyses revealed a main effect of validity, with more positive responses given to valid than invalid arguments, $F(1, 96) = 222.18$, $p < .001$, $\eta_p^2 = .70$. Validity also interacted with reasoning condition, $F(1, 96) = 7.43$, $p = .01$, $\eta_p^2 = .07$. As can be seen in Figure 4.1, individuals in the deduction condition were more influenced by item validity than those in the induction condition. However, sensitivity to argument validity did not differ between OCD groups, $F(1, 96) = 0.16$, $p = .69$, indicating that individuals high in OCD symptoms showed a similar level of discrimination between valid and invalid arguments as those low in OCD symptoms.

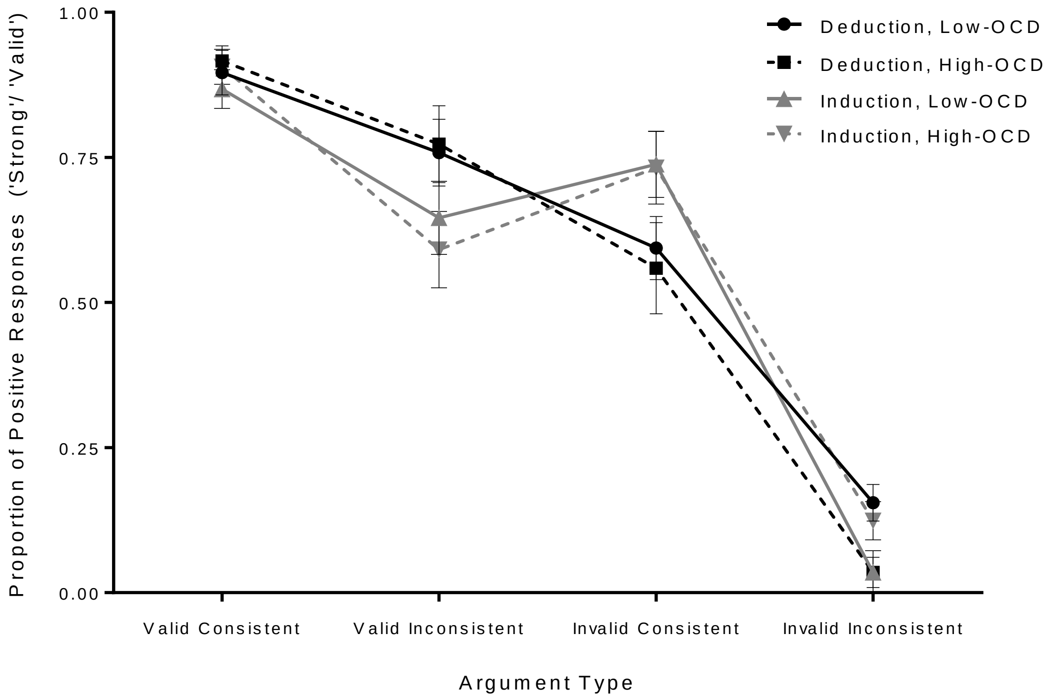

Figure 4.1. Estimated marginal mean proportion of positive ('strong'/ 'valid') responses for each argument type by OCD group and reasoning condition on the Experiment 6 causal consistency induction-deduction task. Error bars represent standard error of the mean.

There was a main effect of causal consistency, with more positive responses given to causally consistent than causally inconsistent arguments, $F(1, 96) = 255.89$, $p < .001$, $\eta_p^2 = .73$. Causal consistency interacted with reasoning condition, $F(1, 96) = 9.80$, $p = .002$, $\eta_p^2 = .09$; as can be seen in Figure 4.1, individuals in the induction condition were more influenced by the causal consistency status of an argument than those in deduction condition. That is, those given induction instructions were more likely to accept causally consistent arguments and reject causally inconsistent arguments than those in the deduction condition.

Validity and causal consistency also interacted, F (1, 96) = 69.89, p < .001, η_p^2 = .42. As can be seen in Figure 4.1, differences in argument acceptance rates between causally consistent and inconsistent arguments were larger for invalid arguments than for valid arguments. However, there was no difference between OCD groups in sensitivity to the causal consistency of an argument, $F(1, 96) = 0.20$, $p = .65$. This indicates that those high and low in OCD symptoms utilised causal consistency information to the same extent in their judgments. No other main effects or interactions reached significance.

To better compare the current results with those of Rips (2001), a follow-up analysis was run that only included the two types of ‘conflict’ arguments: valid/inconsistent and invalid/consistent arguments. This analysis revealed an interaction between reasoning condition and argument type, $F(1, 96) = 10.82$, $p = .001$, $\eta_p^2 = .10$. As in Rips’ (2001) study, under deduction instructions, valid/inconsistent arguments were accepted more often than invalid/consistent arguments, but this pattern reversed under induction instructions. Notably, however, this interaction did not differ between OCD groups, $F(1, 96) = 0.28$, $p = .60$.

All the aforementioned patterns were also found when analyses of positive responding were restricted to those scoring in the upper and lower quartiles of the OCI-R (see Appendix C for details).

Confidence ratings. Estimated marginal mean confidence ratings for argument evaluation judgments for each OCD group, reasoning condition, and argument type are shown in Table 4.2.

Table 4.2

Estimated Marginal Mean Confidence Ratings (out of 5) by OCD Group, Reasoning Condition, Argument Validity and Causal Consistency on the Experiment 6 Causal Consistency Induction-Deduction Task (Standard Deviations in Parentheses)

	Low OCD group		High OCD group	
Argument type	Deduction	Induction	Deduction	Induction
Valid/Consistent	4.32 (0.57)	4.34 (0.54)	4.66 (0.46)	4.26 (0.55)
Valid/Inconsistent	3.99 (0.68)	4.12 (0.74)	4.37 (0.62)	3.86 (0.71)
Invalid/Consistent	3.77 (0.59)	4.11 (0.75)	4.30 (0.56)	3.92 (0.73)
Invalid/Inconsistent	3.97 (0.67)	3.89 (1.01)	4.35 (0.68)	3.84 (0.80)

There was a main effect of validity, such that participants were more confident making judgments about valid (M = 4.23, SD = 0.54) arguments as compared to invalid (M = 4.00, SD = 0.66) arguments, $F(1, 96) = 14.99$, $p < .001$, $\eta_p^2 = .14$. There was also a main effect of causal consistency, such that participants were more confident making judgments about arguments that were causally consistent (M = 4.20, SD = 0.53) as opposed to inconsistent (M = 4.04, SD = 0.65), $F(1, 96) = 9.91$, $p = .002$, $\eta_p^2 = .09$. Furthermore, validity and causal consistency interacted, with the increased confidence for judgments about causally consistent arguments largely driven by valid items, $F(1, 96) = 9.59$, $p = .003$, $\eta_p^2 = .09$, (Valid/Consistent, M = 4.38, SD = 0.55; Valid/Inconsistent, M = 4.08, SD = 0.70; Invalid/Consistent, M = 4.01, SD = 0.68; Invalid/Inconsistent, M = 4.00, SD = 0.82).

There was no main effect of reasoning condition, $F(1, 95) = 2.78$, $p = .10$, nor of OCD group, $F(1, 95) = 1.57$, $p = .21$, on confidence ratings. There was, however, an interaction between OCD group and reasoning condition, $F(1, 95) = 6.87$, $p = .01$, $\eta_p^2 = .07$. Bonferroni-adjusted simple effects analyses with an alpha of .0125 indicated that those high in OCD symptoms were more confident in making judgments of deductive validity (*M* = 4.42, *SD* = 0.47) than inductive plausibility (*M* = 3.97, *SD* = 0.57), $F(1, 96) = 8.10$, $p = .005$, $\eta_p^2 = .08$. In contrast, the confidence of those low in OCD symptoms did not differ between reasoning conditions (deduction, *M* = 4.01, *SD* = 0.46; induction, *M* = 4.12, *SD* = 0.55), $F(1, 96) = 0.64$, $p = .43$. Those high in OCD symptoms were more confident than those low in OCD symptoms in their deductive judgments, $F(1, 96) = 6.67$, $p = .011$, $\eta_p^2 = .06$, but not in their inductive judgments, $F(1, 96) = 0.68$, $p = .41$.

Discussion

This experiment was designed to examine whether individuals high in levels of OCD symptomatology are impaired when using background knowledge (as opposed to inductive sampling heuristics) to make their judgments of inductive plausibility, or when reasoning about arguments where manipulation of argument validity did not depend on class-inclusion. There were two main hypotheses regarding OCD-status, based on previous studies suggesting that inductive reasoning should be impaired in OCD (O'Connor et al., 2005; Pélissier & O'Connor, 2002), and on Experiments 1 and 2, which found that individuals high in OCD symptoms were less sensitive to argument validity than those low in symptoms. It was expected that individuals high in OCD symptoms would be less sensitive to causal consistency and poorer at distinguishing between valid and invalid arguments compared to those low in symptoms.

Overall, participants given deduction instructions were more sensitive to argument validity than those given induction instructions. In contrast, those given induction instructions accepted causally consistent arguments more frequently, regardless of the validity of the argument, compared to those given deduction instructions. This replicates Rips' (2001) finding that people apply different criteria to evaluate arguments when reasoning inductively or deductively (i.e., respectively, consistency with background causal knowledge and logical validity).

Importantly, this response pattern was exhibited by *both* individuals high and low in OCD symptoms. Under induction instructions, both OCD groups were more likely to accept causally consistent arguments as inductively strong, relative to causally inconsistent arguments. This indicates that individuals high in OCD symptoms were just as able to use background causal knowledge to guide judgments of inductive plausibility as those low in symptoms.

Given that all the items used were OCD-neutral in content, it is plausible that any reasoning impairment (R. Kemp et al., 1997; McGuire et al., 2001) or advantage (Johnson-Laird et al., 2006) experienced by individuals high in OCD symptoms may only appear on items with OCD-relevant content. This seems somewhat unlikely, however, as previous studies have generally found inductive impairments in those high in OCD symptoms on tasks that only used OCD-neutral content, or across both OCD-neutral and OCD-relevant content (e.g., Experiment 2; Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007).

In terms of deductive reasoning, both individuals high and low in OCD symptoms accepted valid arguments far more often than invalid arguments, with no reliable group differences in the ability to discriminate between these arguments. This result runs counter to the results of Experiments 1 and 2, where poorer discrimination between valid and invalid items was found in those high in OCD symptomology. This may indicate that those high in OCD symptoms are as adept as those low in symptoms in understanding an argument’s logical structure when validity is not manipulated via class-inclusion relationships.

Alternatively, these results may be accounted for by the structure of Rips' (2001) invalid arguments. Consider the top right-most argument in Table 4.1. Here, Rips argues that although the premise does not logically entail the conclusion (i.e., the conclusion is not *necessarily* true), the probability of the conclusion is high due to its consistency with background knowledge. However, Oaksford and Hahn (2007) argue that on a propositional level, these invalid arguments actually have no logical structure at all. To rectify this, and to ensure that invalid arguments are directly comparable to valid ones, Oaksford and Hahn argue that a major premise must be inserted in the invalid arguments. This might be accomplished by using a different invalid problem structure, such as affirming the consequent (if *p* then *q*, *q* therefore *p*). For example, the top right-most argument of Table 4.1 now becomes:

If Car #10 runs into a brick wall, it will stop

Car #10 stops

Car #10 runs into a brick wall

A further issue with the current item structure is that participants could correctly discriminate between valid arguments containing two premises and invalid arguments that contained only one premise, once this pattern was recognised. After the first few trials, participants could have made their judgments based on the *number* of premises presented (valid arguments contained two premises and invalid arguments contained one) without consideration of the logical structure of the argument. Although there is evidence to suggest that this does not completely account for the data (e.g., invalid/consistent arguments were endorsed by those in the deduction condition 58% of the time on average), this possibility cannot be excluded with the current argument structure. Experiment 7 addressed this issue by re-examining deductive reasoning in high and low OCD symptom groups on valid and invalid arguments that were more comparable in length and structure.

Overall, individuals high in OCD symptoms were just as confident in their argument evaluation decisions as those low in OCD symptoms. However, the two groups showed different patterns of relative confidence in inductive and deductive judgments. Whilst the confidence of individuals low in OCD symptoms did not differ across reasoning condition, those high in symptoms were more confident making judgments of deductive validity than inductive plausibility. Individuals high in OCD symptoms were more confident than those low in OCD symptoms when making deductive, but not inductive, judgments. This is broadly consistent with a “spared deduction, impaired induction” pattern. Contrary to the hypothesis suggested by O'Connor et al. (2005), however, the obtained pattern reflects differences in relative confidence in deductive and inductive reasoning, rather than differences in objective reasoning performance.

Experiment 7

Dysfunctional beliefs are hypothesised to form a causal basis for many catastrophic misappraisals present in OCD (Salkovskis, 1999), as well as in a range of other anxiety disorders (e.g., panic disorder, social anxiety disorder, and post-traumatic stress disorder; D. M. Clark, 1986; D. M. Clark & Wells, 1995; Ehlers & Clark, 2000). These models suggest that dysfunctional beliefs cause an individual to either interpret danger where none exists or to overemphasise the catastrophic nature of a situation. By way of example, many OCD patients believe that experiencing an unwanted and unacceptable intrusive thought makes the event more likely to occur, in what is known as thought-action fusion (see Shafran & Rachman, 2004 for a review and discussion of other types of thought-action fusion). If one believes that having a thought is equivalent to performing the content of the thought (Salkovskis, 1985), then the intrusion ‘I might run someone over’ is understandably distressing, and is likely to prompt neutralising urges (e.g., checking the rear view mirror repeatedly until some subjective standard of certainty is met).

If dysfunctional beliefs are crucial to the aetiology of OCD, then it is vital to understand the persistence of these beliefs in the face of disconfirming evidence. Why might an OCD patient continue to believe that the intrusion ‘I might have run someone over’ signifies that they are likely to have committed the action, despite numerous instances where this has not been the case? One explanation may be the role of reasoning in the maintenance of dysfunctional beliefs.

Specifically, belief bias, the tendency to endorse conclusions that are consistent with one’s beliefs about the world and to reject conclusions that are inconsistent with one’s beliefs, regardless of validity (e.g., Evans et al., 1983; Evans, Handley, & Harper, 2001; Evans, Newstead, & Byrne, 1993; Evans, Over, et al., 1993; Newstead & Evans, 1993), may play a role in the maintenance of OCD-specific beliefs. In situations where time and capacity are limited, and where logically valid conclusions are not vital, Evans and colleagues argue that relying on previously held beliefs is adaptive for generating likely conclusions. However, if beliefs are dysfunctional, then this is a counterproductive strategy, likely to impede the disconfirmation of these beliefs (de Jong, Weertman, Horselenberg, & van den Hout, 1997; Smeets & De Jong, 2005).

A task specifically designed to examine the ability to discriminate between valid and invalid arguments in circumstances where argument conclusions are either consistent (i.e., believable) or inconsistent (i.e., unbelievable) with existing beliefs is known as the ‘belief bias’ task. Belief bias is typically studied using ‘categorical syllogisms'[9], which contain two premises and a conclusion. These are constructed with three terms: the predicate (X), which is the non-repeated term of the first premise; the middle term (Y); and the subject (Z), which is the non-repeated term of the second premise (Dube, Rotello, & Heit, 2010). Conclusions are generated through the linking of the predicate and the subject via the middle term. For example, a valid syllogism (where the conclusion necessarily follows from the premises) could have the following structure:

All X are Y

No Y are Z

No Z are X

When these abstract terms are replaced with meaningful content (see Table 4.3 for examples), conclusions can also be perceived as believable or unbelievable in light of background knowledge. This allows for an examination of an individual's ability to correctly follow the logical structure of an argument under conditions where the conclusion does or does not align with one's existing beliefs (i.e., an examination of belief bias).

Note that in the wider reasoning literature, two sources of belief bias have been identified: 1) a general response bias to endorse believable conclusions rather than unbelievable ones, regardless of logical validity (hereafter referred to as the 'belief response bias'), and 2) a belief by validity interaction, where there is better discrimination between valid and invalid arguments when conclusions are unbelievable than when they are believable (Evans et al., 2001; Klauer, Musch, & Naumer, 2000).

[9] Many of the arguments used in the earlier experiments also contained categorical content. However, the term 'categorical syllogism' has been commonly used in the wider reasoning literature to refer to a particular argument structure which uses quantifiers such as 'All', 'No' or 'Some'. This terminology is followed in this thesis.

One study to examine belief bias in OCD noted that OCD patients were poorer at distinguishing between valid and invalid syllogisms when item content was OCD-neutral, relative to non-clinical controls (Simpson et al., 2007). This effect was primarily driven by OCD patients' inability to accept valid OCD-neutral arguments that featured unbelievable conclusions (OCD patients accepted 42% of these arguments, relative to the 67% that non-clinical controls accepted). An interaction between believability and validity was only reliable for OCD patients, not controls, and was larger for OCD-neutral items.

It is notable that these differences were restricted to (or were larger for) OCD-neutral items. This pattern is consistent with Johnson-Laird et al.'s (2006) hypothesis that familiarity with OCD-relevant content should facilitate logical reasoning involving such content. Relatedly, Simpson et al. (2007) suggest that OCD patients perform poorly on OCD-neutral items relative to non-clinical controls because they lack confidence when reasoning about content that is not part of their everyday concern. These feelings of uncertainty are argued to activate heuristic responding, which, in the case of belief bias, is to respond according to the believability of an argument, rather than its logical validity (Quayle & Ball, 2000).

In line with Simpson et al. (2007), it was expected that individuals high in OCD symptoms would be less able to discriminate between valid and invalid arguments (at least on OCD-neutral items). Such a finding would parallel the results of Experiments 1 and 2, indicating that those high in OCD symptomatology are impaired in their deductive reasoning. Importantly, as the syllogisms in Experiment 7 were not subject to the same limitations as those used in Experiment 6 (e.g., valid and invalid syllogisms were of equal length), finding a deductive impairment for the high OCD group would suggest that the structure of the arguments in Experiment 6 facilitated discrimination between valid and invalid arguments.

Further, based on Simpson et al.'s (2007) suggestion that feelings of uncertainty when reasoning about unfamiliar content activates greater heuristic responding, it was also expected that individuals high in OCD symptoms would exhibit larger belief biases relative to those low in symptoms. That is, individuals high in OCD symptoms were expected to be more influenced by the believability of an argument than those low in symptoms. Notably, this study included belief bias items containing either OCD-neutral or OCD-related content, as Simpson et al.’s results suggest that group differences in belief bias may be accentuated when reasoning about OCD-neutral items. Specifically, those high in OCD symptoms were expected to exhibit a larger belief by validity interaction than those low in OCD symptoms (i.e., discrimination between valid and invalid items was expected to be more affected by believability in those high in OCD symptoms). As previous experiments noted group differences after controlling for general ability (Experiments 2, 3 and 5), it was also expected that these hypothesised impairments would be robust over and above any group differences in general ability.

Method

Participants

One hundred and seventeen undergraduate students enrolled in an introductory Psychology course undertook this experiment in return for course credit. These participants also took part in Experiment 4. In line with the previous experiments, 13 participants were excluded because English was not their primary language. For the purpose of analysis, these participants were divided into two groups on the basis of their OCI-R scores. Fifty participants scored 21 or higher on the OCI-R, and were classified as the high OCD group (*M* = 32.22, *SD* = 8.05); of these participants, 29 participants were female and the mean age of this group was 18.92 years (*SD* = 8.04). There were 54 participants that scored less than 21 on the OCI-R (*M* = 9.56, *SD* = 5.71); of these, 26 were female and the mean age of this group was 20.39 (*SD* = 5.12). The two groups did not differ by age, $F(1, 102) = 3.67$, $p = .06$, nor in the distribution of gender across groups, $\chi^2(1, N=104) = 1.01$, $p = .33$.

Materials and Procedure

Participants undertook the computerised experiment individually in a quiet room. Half the participants began the experiment with the belief bias task, before completing the premise-diversity hypothesis testing task (see Experiment 4, Chapter 3). The remaining participants completed the tasks in reversed order. Following this, all participants completed the OCI-R. All participants were administered the two-subscale version of the WASI by the author to complete the experiment. The belief bias task was the only new task, and was created using the programming package Livecode©.

Thirty-two syllogisms were constructed for the belief bias task. Each syllogism contained two premises presented above a line and a conclusion presented below the line (see Table 4.3 for examples and Appendix C for a complete list of stimuli). Arguments varied by validity and believability. Half of the syllogism frames were deductively valid, whilst the remaining frames were deductively invalid. Half the conclusions presented were believable (e.g., ‘some used first aid products are not used bandaids’), and the remaining conclusions were unbelievable (e.g., ‘some used bandaids are not used first aid products’). Believability status was altered by reversing the order of assignment of terms to subject and predicate positions. Thus, the 32 syllogisms were divided equally into four syllogism types (valid/believable, valid/unbelievable, invalid/believable and invalid/unbelievable).

Table 4.3

Examples of One Syllogistic Frame by OCD-Relevance and Argument Type on the Experiment 7 Belief Bias Task

	OCD-relevance	
Argument type	Neutral	Relevant
Valid/Believable	No oaks are trivels Some trees are trivels ________________ *Some trees are not oaks*	No plugged in toasters are bictoids Some plugged in appliances are bictoids ________________ *Some plugged in appliances are plugged in toasters*
Valid/Unbelievable	No trees are trivels Some oaks are trivels ________________ *Some oaks are not trees*	No plugged in appliances are bictoids Some plugged in toasters are bictoids ________________ *Some plugged in toasters are not plugged in appliances*
Invalid/Believable	No trees are trivels Some oaks are trivels ________________ *Some trees are not oaks*	No plugged in appliances are bictoids Some plugged in toasters are bictoids ________________ *Some plugged in appliances are not plugged in toasters*
Invalid/Unbelievable	No oaks are trivels Some trees are trivels ________________ *Some oaks are not trees*	No plugged in toasters are bictoids Some plugged in appliances are bictoids ________________ *Some plugged in toasters are not plugged in appliances*

Note. The premises are given in normal font above the line and are assumed to be true. Conclusions are given in italics below the line. An argument is ‘valid’ if necessarily true.

The syllogisms were also divided by OCD-relevance, such that half the arguments featured OCD-relevant content generated by the author, whilst the remaining arguments featured neutral content taken from Dube, Rotello and Heit’s (2010) study (e.g., content related to animals, trees, and transportation). The subject terms used were invented non-words (e.g., ‘bictoid’, ‘junaric’), in order to minimise the effect of premise believability.

Further, the item set was composed of eight syllogistic problem frames (see Appendix C) that controlled for three factors that contribute to syllogism difficulty: atmosphere (the drawing of conclusions based on the global impression produced by the premises, e.g., whether the premises are affirmative or negative, universal or particular; Beggs & Denny, 1969), conversion (the erroneous acceptance of the converse of universal affirmative or particular negative statements, when this is not logically permissable; Dickstein, 1975; Dickstein, 1981; Revlin, Leirer, Yopp, & Yopp, 1980), and figural effects (variances in the order in which terms occur in the major and minor premises; Dickstein, 1978). Four of these problem frames controlled for atmosphere, conversion and figural effects, whilst the remaining four problems controlled for atmosphere and figure but not conversion. Atmosphere and conversion were controlled by using both valid and invalid problem forms, with the logically convertible quantifiers ‘some’ and ‘no’ paired with the conclusion quantifier ‘some... are not’; this is favoured by the atmosphere created by ‘some’ and ‘no’ (Dube, 2009). These problem structures do not allow for illicit conversion. Figure was controlled for by presenting conclusions in both middle-predicate and predicate-middle directions (Dube, 2009).

To begin the task, and consistent with the procedures used in the previous experiments, participants were given definitions of ‘valid’ (the conclusion is *necessarily true* if the premises are true) and ‘invalid’ (the conclusion is *not necessarily true* if the premises are true) arguments. For each item, participants made binary judgments about the validity of each item. Untimed responses were made via mouse click. Syllogisms remained on the screen until participants made a response. Following each decision, participants rated their confidence in their decision on a Likert scale from 1 (not confident at all) to 5 (very confident).

Results

These analyses examined the proportion of arguments that were endorsed as 'valid'. Higher endorsement rates were equivalent to more accurate responding on valid items, but less accurate responding on invalid items. Measures of belief bias were also examined, with comparisons drawn between individuals high and low in OCD symptoms (clinical cut-off and quartile analyses). Following this, confidence ratings were examined (see Appendix C for full analyses). Ten participants were excluded from data analyses, as they endorsed 75% or more of the invalid/unbelievable items, and were therefore assumed not to have understood the experimental instructions. Eight of these participants were from the low OCD symptom group. None of the reported results change if these participants are included in the analyses.

Preliminary Analyses

ANOVA analyses confirmed that there was no effect of task order (belief bias task or premise diversity hypothesis-testing task first) on judgments of item validity, $F(1, 89) = 2.87$, $p = .09$. Therefore, all subsequent analyses were collapsed across task order. A one-way ANOVA analysis of general ability scores between individuals high and low on OCD symptoms indicated that those high in OCD symptoms had lower general ability estimates ($M = 101.46$, $SD = 1.30$) relative to those low in symptoms ($M = 108.94$, $SD = 1.33$), $F(1, 92) = 16.21$, $p < .001$, $\eta_p^2 = .15$.

Belief Bias Task

Initial analyses examined the proportion of arguments endorsed as 'valid'. Estimated marginal means for these items can be seen in Figure 4.2. The proportion of arguments evaluated as valid was analysed in a 2 (level of OCD-symptoms: high or low) × (2) (item-validity: valid or invalid) × (2) (conclusion-believability: believable or unbelievable) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA analysis with repeated measures on the last three factors and controlling for general ability.

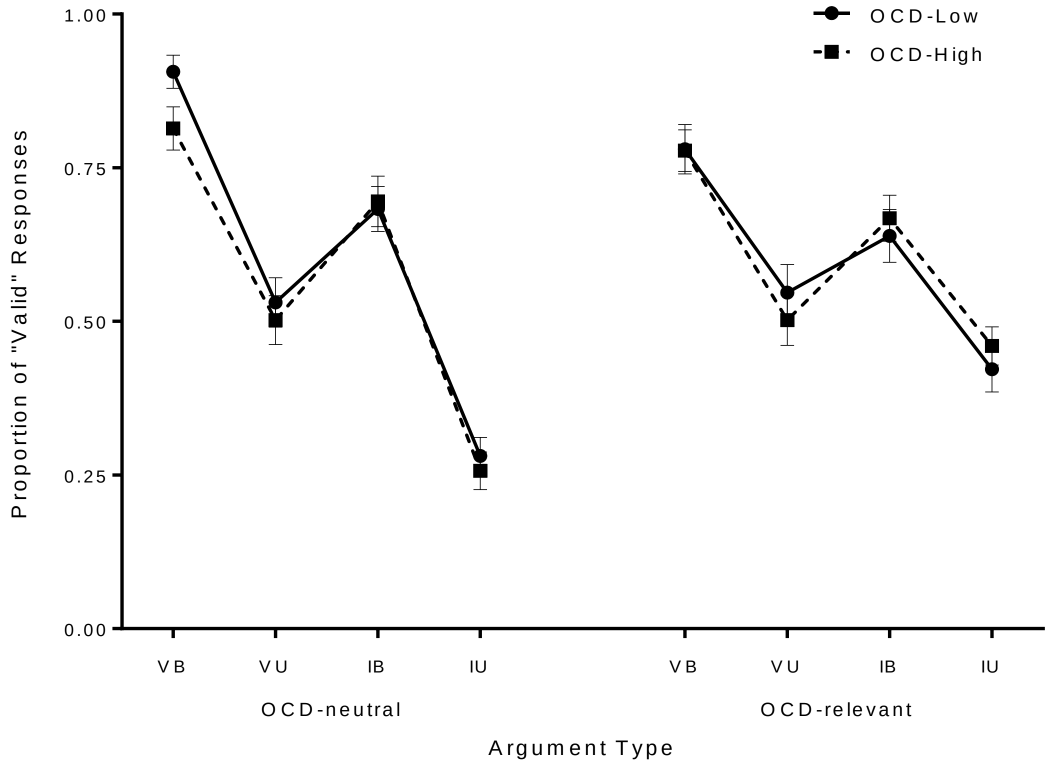

Figure 4.2. Estimated marginal mean proportion of arguments evaluated as 'valid' by argument type and OCD group on the Experiment 7 belief bias task. VB = valid/believable, VU = valid/unbelievable, IB = invalid/believable, IU = invalid/unbelievable. Error bars represent standard error of the mean.

There was a main effect of validity, such that logically valid arguments were more likely to be judged as valid than invalid arguments, $F(1, 92) = 64.29$, $p < .001$, $\eta_p^2 = .41$. Item validity also interacted with OCD-relevance, $F(1, 92) = 11.95$, $p = .001$, $\eta_p^2 = .11$, such that participants were better at discriminating between valid and invalid arguments when item content was OCD-neutral, as can be seen in Figure 4.2. Importantly, validity and OCD group did not interact, $F(1, 92) = 2.04$, $p = .16$, indicating that both high and low OCD groups were equally able to distinguish between valid and invalid items.

There was also a main effect of believability, such that arguments with believable conclusions were more likely to be judged as valid than arguments with unbelievable conclusions, $F(1, 92) = 178.24$, $p < .001$, $\eta_p^2 = .66$. That is, there was a ‘belief response bias’ effect. This effect did not interact with OCD group, $F(1, 92) = 0.00$, $p = .97$, indicating that both groups exhibited a similar preference for endorsing believable over unbelievable arguments. However, believability did interact with OCD-relevance, $F(1, 92) = 29.06$, $p < .001$, $\eta_p^2 = .24$. As shown in Figure 4.2, the tendency to accept believable conclusions over unbelievable ones (regardless of logical validity) was larger for OCD-neutral items than for OCD-relevant items.

There was a 3-way interaction between validity, believability and OCD-relevance, $F(1, 92) = 4.35$, $p = .04$, $\eta_p^2 = .05$. Inspection of Figure 4.2 and follow-up tests showed that for OCD-neutral items, there was better discrimination between valid and invalid arguments when these featured unbelievable (valid/unbelievable vs. invalid/unbelievable; $M_{VU\text{-}IU} = 0.25$) than believable conclusions (valid/believable vs. invalid/believable; $M_{VB\text{-}IB} = 0.17$), $F(1, 92) = 4.34$, $p = .04$, $\eta_p^2 = .05$. This pattern, however, disappeared for OCD-relevant items (unbelievable items, $M_{VU\text{-}IU} = 0.08$; believable items, $M_{VB\text{-}IB} = 0.13$), $F(1, 92) = 0.85$, $p = .36$. That is, the ‘belief bias interaction’ effect was only noted on OCD-neutral items. No other main effects or interactions, including all those involving OCD symptom status, were significant. The results were unchanged when group comparisons involved individuals scoring highest and lowest in OCD symptoms (see Appendix C for more detail).

Confidence ratings. The mean confidence ratings for each OCD group and argument type are shown in Table 4.4. Participants were more confident in their evaluation of OCD-neutral (*M* = 3.77, *SD* = 0.63) than OCD-relevant (*M* = 3.45, *SD* = 0.68) items, $F(1, 92) = 79.00$, $p < .001$, $\eta_p^2 = .46$. They were also more confident in evaluating valid (*M* = 3.69, *SD* = 0.66) than invalid (*M* = 3.53, *SD* = 0.64) arguments, $F(1, 92) = 23.05$, $p < .001$, $\eta_p^2 = .20$, and in evaluating believable (*M* = 3.70, *SD* = 0.66) than unbelievable (*M* = 3.53, *SD* = 0.66) arguments, $F(1, 92) = 15.69$, $p < .001$, $\eta_p^2 = .15$. Furthermore, validity and believability interacted (Valid/Believable, *M* = 3.81, *SD* = 0.71; Valid/Unbelievable, *M* = 3.57, *SD* = 0.70; Invalid/Believable, *M* = 3.58, *SD* = 0.68; Invalid/Unbelievable, *M* = 3.48, *SD* = 0.70), $F(1, 92) = 5.76$, $p = .02$, $\eta_p^2 = .06$. As shown in Table 4.4, the effects of believability on confidence were more marked for valid than for invalid arguments. There was no main effect of OCD group, $F(1, 91) = 0.86$, $p = .36$. No other main or interaction effects reached significance.

Table 4.4

Estimated Marginal Mean Confidence Ratings (out of 5) by OCD Group, OCD-Relevance, and Argument Type on the Experiment 7 Belief Bias Task (Standard Deviations in Parentheses)

	OCD-neutral		OCD-relevant	
Argument type	Low OCD	High OCD	Low OCD	High OCD
Valid/Believable	4.12 (0.76)	3.88 (0.72)	3.70 (0.88)	3.55 (0.75)
Valid/Unbelievable	3.80 (0.82)	3.63 (0.72)	3.54 (0.76)	3.33 (0.74)
Invalid/Believable	3.74 (0.72)	3.80 (0.74)	3.42 (0.83)	3.36 (0.76)
Invalid/Unbelievable	3.67 (0.79)	3.55 (0.65)	3.44 (0.89)	3.28 (0.69)

Discussion

This experiment re-examined whether individuals high in OCD symptoms have difficulty distinguishing between valid and invalid arguments relative to those in low OCD symptoms, particularly when the logical validity of an argument conflicted with the believability of the conclusion. Contrary to expectations, no differences were found between OCD groups on either sensitivity to argument validity or in the tendency to endorse believable conclusions over unbelievable ones (‘belief response bias’). This was the case for both OCD-neutral and OCD-relevant items. Further, although for OCD-neutral items, discrimination between valid and invalid arguments was better when the conclusion was unbelievable (i.e., ‘belief bias interaction’), this effect was found in both high and low OCD symptom groups.

Thus, individuals high and low in OCD symptoms seemed equivalent in their ability to differentiate between valid and invalid arguments. This finding differs from previous work demonstrating that those high in OCD symptoms were impaired in their understanding of argument validity (e.g., Experiments 1 and 2; Simpson et al., 2007). In the current experiment, however, the manipulation of argument validity was not dependent on class-inclusion arguments as it was in Experiments 1 and 2.

The current design was similar to that used by Simpson et al. (2007). However, where Simpson and colleagues found that OCD patients accepted equal numbers of valid and invalid OCD-neutral items and exhibited a larger belief bias on these items, the current experiment did not. Simpson et al. (2007) suggested that OCD patients’ impairments may be explained by their feelings of uncertainty when reasoning with content that is not part of their everyday concern (i.e., OCD-neutral content). These feelings of uncertainty are argued to prompt a heuristic response (i.e., responding according to believability).

One important methodological difference between the two studies is that the current experiment used nonsensical premise content (e.g., ‘no plugged in toasters are bictoids’), whereas Simpson et al. used meaningful content (e.g., ‘no cancers are detectable’). Removal of meaningful premise content reduces the influence of premise believability, which has been noted to influence the extent to which people are willing to accept conclusions (Stevenson & Over, 1995; Thompson, 1996). Thus, it may be that removing meaningful premise content provided an additional cue to the logical structure of the arguments, increasing the ease and confidence with which individuals high in OCD symptoms could make deductive inferences about OCD-neutral items. This would imply that deficits in deductive reasoning on categorical syllogisms are only apparent when the logical structure of the arguments is not so clear (e.g., where the premises and conclusions contain familiar content).

Finally, in line with Experiment 2 (but in contrast to Experiment 1), there was no overall group difference in confidence. Thus, despite the conventional conceptualisation of OCD as a ‘doubting’ disorder, those high in OCD symptomatology do not always appear less confident when drawing conclusions from given evidence than those low in symptomatology. This is in contrast to the finding that those high in OCD symptoms were more confident than those low in symptoms when reasoning deductively (Experiment 6). This may be related to the complexity of the items in the current experiment. Overall, participants were more confident in Experiment 5 (M = 4.19, SD = 0.50) than in Experiment 6 (M = 3.77, SD = 0.63) when reasoning about OCD-neutral items under deduction instructions; this seems to indicate that the current item set was more complex. Thus, it may be the case that those high in OCD symptoms only exhibit elevated confidence when making simpler deductions, but not when reasoning about more complex syllogisms.

Summary

Collectively, Experiments 6 and 7 imply that individuals high in OCD symptoms are just as able as individuals low in symptoms to distinguish between valid and invalid arguments when the logical structure of these arguments are manipulated by a form other than class-inclusion (i.e., modus ponens, conjunctive syllogisms or categorical syllogisms). These high OCD symptom individuals may be more confident than low symptom individuals when reasoning about simple deductive problems. Further, those high in OCD symptoms appear as able to employ background knowledge in their judgments of inductive plausibility as are those low in symptoms.

CHAPTER 5 - Deductive and Inductive Reasoning in a Clinically Diagnosed Sample

The preceding experiments have demonstrated that the “spared deduction, impaired induction” account of OCD (O'Connor et al., 2005; Pélissier & O'Connor, 2002) may not apply to all forms of deductive and inductive reasoning. Individuals high in OCD symptoms demonstrated a reduced sensitivity to deductive validity when validity was manipulated via class-inclusion (Experiments 1 and 2) but not via modus ponens, conjunctive syllogisms, or categorical syllogisms (Experiments 6 and 7), relative to those low in symptoms. Those high in OCD symptoms appear unimpaired in their use of sample size (Experiment 1) and background knowledge (Experiment 6) as guides to judgments of inductive plausibility. However, these individuals seem to exhibit a reduced preference for diverse premise pairs (Experiments 2 and 3) relative to individuals low in OCD symptoms. This may be partially explained by differences in the way that individuals high in OCD symptomatology perceive similarity of premise pairs relative to individuals low in symptomatology (Experiment 5).

So far, all experiments have made use of analogue samples high or low in OCD symptomology. As noted in Chapter 1, this approach is justified because OCD is a dimensional disorder (D. A. Clark, 2005; Mataix-Cols et al., 2005), with analogue samples high in OCD symptoms exhibiting the same type of symptomatology as OCD patients, differing only in severity (Abramowitz et al., 2014; Gibbs, 1996).

Nevertheless, if the reasoning deficits noted thus far have implications for the aetiology and/or maintenance of OCD symptoms, then these deficits should also be found in a clinically diagnosed sample. This is especially important as the previous studies examining deductive and inductive reasoning in OCD have used clinical samples (e.g., Aardema et al., 2009; Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007).

Experiment 8

The current experiment examined deductive and inductive reasoning in a clinically diagnosed sample of OCD patients. The reasoning tasks used were those from the previous experiments where deficits were found in an analogue sample high in OCD symptomatology. Specifically, the current experiment examined OCD patients' sensitivity to argument validity on the basis of class-inclusion on a premise monotonicity induction-deduction task (cf. Experiment 1), their sensitivity to premise diversity on a forced-choice task (cf. Experiment 3), and their perception of the similarity of the premise pairs in the premise diversity task (cf. Experiment 5), in comparison to a group of non-clinical controls.

In line with the performance of the analogue samples high in OCD symptoms, it was expected that relative to non-clinical controls, OCD patients would be less sensitive to argument validity on a class-inclusion task and less sensitive to premise diversity on a property induction task. Specifically, OCD patients were expected to show less discrimination between valid and invalid arguments, and to exhibit a reduced preference for diverse premises (both on forced-choice judgments and argument strength ratings) as a basis for inductive generalisation, relative to controls. Based on the outcomes of the analogue sample studies, group differences in reasoning were expected to be general across both OCD-neutral and OCD-relevant content, and to persist after group differences in general ability were controlled.

Following Experiment 5, the possibility of group differences in the perception of the relative similarity of argument premises was also examined. Based on the analogue sample results, it was expected that OCD patients would be less likely to endorse the previously classified similar/non-diverse items as the most similar option to a given premise, compared to non-clinical controls.

Method

Participants[10]

Twenty-six participants that met criteria for a clinical diagnosis of OCD were recruited via advertisements placed in local newspapers and online. Participants were diagnosed according to the DSM-5 (American Psychiatric Association, 2013) criteria. Participants were excluded if they exhibited any of the following: past or current drug or alcohol abuse, traumatic brain injury, suicidality, or symptoms of psychosis. Most participants (*N* = 23) scored above 20 (the suggested cut-off for discriminating between patients with OCD and non-clinical controls) on the OCI-R. Two participants had OCI-R scores below 20, but were considered suitable for inclusion because they met diagnostic criteria for OCD. The other participant was excluded as English was not her primary language.

Only 20% of these 25 participants met criteria for OCD alone. Many had one or more co-morbid diagnoses. These included generalised anxiety disorder (60%), social anxiety disorder (48%), specific phobia (40%), major depressive disorder (32%), separation anxiety disorder (28%), panic disorder (20%), illness anxiety (20%), and post-traumatic stress disorder (12%). Scores on the OCI-R (see Table 5.1) for the OCD sample were comparable to scores from samples of OCD patients in previous studies (e.g., Abramowitz, Tolin, & Diefenbach, 2005; Foa et al., 2002; Lorian & Grisham, 2011; Storch et al., 2008), and were higher than scores from the majority of analogue samples employed in Experiments 1 to 7 (see Appendix C for comparisons). A

[10]An anxious control group was also recruited (i.e., participants who met DSM-5 diagnostic criteria for an anxiety disorder but did not meet criteria for OCD). However, only a small sample (*N* = 14) was found who satisfied these criteria, and therefore, the anxious controls were not included in the main analysis. See Appendix D for relevant task means.

substantial proportion (32%) of the OCD patients were taking medication for their anxiety or depressive difficulties.

Of the 25 OCD participants, 36% graduated high school as their highest level of educational training, 40% had an undergraduate degree, 8% had a Masters degree and the remaining 16% had trade or vocational training. Furthermore, 36% of patients indicated they were of Asian descent, 52% were Caucasian, and the remaining 12% indicated they were of neither Asian, Caucasian, or Hispanic descent. Seventeen of these participants were female.

Non-clinical control participants were recruited via advertisements in local newspapers and online. These participants were screened for current mental health problems and any history of substance abuse or traumatic brain injury. Twenty-seven non-clinical controls were recruited; two were excluded as English was not their primary language. Of the remaining 25 participants, 12% were high school graduates, 40% had an undergraduate degree, 8% had a doctorate degree, 28% had a Masters degree, and the remaining 12% had trade or vocational training. Furthermore, 64% of the non-clinical controls indicated they were of Asian descent, 28% were Caucasian, 4% were of Hispanic descent, and the remaining 4% were of a different descent. Fifteen of these participants were female.

There were no differences in the distribution of the OCD patients and non-clinical controls across highest education level, $\chi^2(4, N=50) = 7.92$, $p = .10$, ethnicity, $\chi^2(3, N=50) = 5.76$, $p = .12$, or gender, $\chi^2(2, N=50) = 0.35$, $p = .77$. The remaining group demographics are displayed in Table 5.1. All participants received $20 per hour for their participation.

Table 5.1

Mean Values and t-test Comparisons for Clinical Group Demographics (Standard Deviations Shown in Parentheses and Statistically Reliable Differences between Groups Shown in the Final Column)

	Group				
Measures	OCD patient	Non-Clinical Control	*t*	*df*	*p*
Age (years)	32.76 (15.41)	35.52 (11.32)	.72	48	.47
General ability	112.72 (9.45)	109.12 (12.89)	1.13	48	.27
OCI-R score	37.84 (12.21)	5.52 (5.21)	12.17	48	<.001**
Depression score	10.60 (5.90)	1.56 (2.87)	6.89	48	<.001**
Anxiety score	7.84 (4.47)	0.88 (1.39)	7.43	48	<.001**
Stress score	11.08 (5.65)	2.72 (3.34)	6.37	48	<.001**

Note. **p* < .05, ***p* < .01.

Materials and Procedure

At the beginning of the experimental session, the Anxiety Disorder Interview Schedule-5 (ADIS-5; Brown & Barlow, 2014) was administered by either the author or another clinical psychology intern. Both investigators were in the final year of a postgraduate degree in Clinical Psychology and held provisional accreditation with the Australian Psychology Accreditation Council. Interviews took approximately 20 minutes for non-clinical controls and 40 minutes for OCD participants.

The computerised section of the experiment was created using Livecode©, and included the collection of demographic information, questionnaires (OCI-R and DASS-21), and three reasoning tasks (premise monotonicity, premise diversity and similarity ratings). Participants completed this individually in a quiet testing room. Following the completion of the computerised section, the WASI was administered by the experimenter. A detailed listing of task order is shown in Table 5.2. The majority of participants took approximately 2.5 to 3 hours to complete the entire experiment.

Table 5.2

Order of Task Administration in Experiment 8

Order	Task
1	ADIS-5 (experimenter administered). Followed by a 5-minute break.
2	Demographic information
3	OCI-R
4	Premise monotonicity (induction or deduction instructions)
5	Diversity task. Followed by a 10-minute break.
6	DASS-21
7	Premise monotonicity (induction or deduction instructions)
8	Similarity rating task. Followed by a 5-minute break.
9	WASI (experimenter administered)
10	Debrief

Premise monotonicity induction-deduction task. The premise monotonicity task was similar to the corresponding task used in Experiment 1, but was modified as follows. First, only two argument lengths were used (one-premise and five-premise arguments). Second, reasoning condition was manipulated within-subjects, such that all participants evaluated arguments under both deduction and induction instructions. Administration of the deduction and induction blocks was separated by other tasks (see Table 5.2). Reasoning block order was counterbalanced across participants. Lastly, unlike Experiment 1 (but in line with Experiments 2-5 and 7), OCD-relevant items as well as OCD-neutral items were used (see rows 3 and 4 in Table 5.3 for examples, and Appendix D for a full list of stimuli).

Table 5.3

Examples of the Six Argument Types Used in the Experiment 8 Premise Monotonicity Induction-Deduction Task

OCD-relevance	Argument length	Argument validity	
		Valid	Invalid
Neutral	One-premise	All mammals have Property X ———————— *All horses have Property X*	All donkeys have Property X ———————— *All horses have Property X*
	Five-premise	All mammals have Property X All wolves have Property X All goats have Property X All monkeys have Property X All dogs have Property X ———————— *All horses have Property X*	All donkeys have Property X All wolves have Property X All goats have Property X All monkeys have Property X All dogs have Property X ———————— *All horses have Property X*
Relevant	One-premise	All switched on household appliances have Property Y ———————— *All switched on blenders have Property Y*	All switched on irons have Property Y ———————— *All switched on blenders have Property Y*
	Five-premise	All switched on household appliances have Property Y All switched on toasters have Property Y All switched on microwaves have Property Y All switched on washing machines have Property Y All switched on irons have Property Y ———————— *All switched on blenders have Property Y*	All switched on electric kettles have Property Y All switched on toasters have Property Y All switched on microwaves have Property Y All switched on washing machines have Property Y All switched on irons have Property Y ———————— *All switched on blenders have Property Y*

Note. The premises are given in normal font above the line and are assumed to be true. Conclusions are given in italics below the line. An argument is ‘strong’ if the conclusion is seen as plausible given the premises, and is ‘valid’ if necessarily true given the premises.

Aside from these modifications, the task remained unchanged, with participants making binary judgments about either the validity (deduction instructions) or plausibility (induction instructions) of argument conclusions based on the given premises. After each judgment, participants rated their confidence in their decision on a scale ranging from 1 (not confident at all) to 5 (very confident).

An item set containing 80 arguments was created, with items split evenly between the factors of OCD-relevance, validity, and length. That is, there were 10 of each argument type (valid one-premise, valid five-premise, invalid one-premise, and invalid five-premise items, crossed by OCD-relevance). Similar to Experiment 1, a quarter of the invalid arguments featured general conclusions, whereas the remaining arguments featured specific conclusions. Individual items were randomly allocated to either deduction or induction blocks, with the constraint that participants saw five of each argument type in each version of the task (i.e., a total of 40 items per block). No time limits were placed on any of the judgments, and all responses were made via mouse click.

Forced-choice premise diversity task. This task was identical to that used in Experiment 3. Participants made 30 forced-choice decisions regarding whether diverse or non-diverse premise pairs were stronger evidence for a given claim (conclusion). The item set was unchanged, with items divided equally between OCD-neutral and OCD-relevant content. Item order was randomised for each participant. After each binary judgment, participants rated the usefulness of both the diverse and non-diverse arguments in evaluating the claim on a scale from 1 (not useful at all) to 10 (very useful). Judgments were not time limited, and responses were made via mouse click.

Similarity ratings task. This task was the same as that used in Experiment 5, except that it was presented using Livecode©, rather than online survey software.

Participants made binary decisions about which of two premise pairs were more similar in content via mouse click; these 30 premise pairs were the premises presented in the premise-diversity task, without the accompanying conclusions. Participants then rated the similarity of each pair of premises on a Likert-type scale from 1 (not at all similar) to 10 (extremely similar).

Depression Anxiety Stress Scales - 21 item version. The Depression Anxiety Stress Scales (DASS-21; Lovibond & Lovibond, 1995) were designed to measure an individual's depression, anxiety, and stress levels over the past week on a 4-point severity/frequency scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much, or most of the time). Examples of items include: "I felt that I had nothing to look forward to" (depression), "I felt I was close to panic" (anxiety), and "I found it hard to wind down" (stress). Scores are determined by summing each relevant depression, anxiety and stress item (see Appendix D for the full questionnaire). The DASS-21 has high internal consistency and concurrent validity (Antony & Swinson, 1998).

Results

Preliminary Analyses

A one-way ANOVA analysis of general ability (WASI) scores between OCD patients and non-clinical controls indicated that the groups did not differ in estimates of general ability, with OCD patients achieving a mean general ability score of 112.72 (SD = 2.26) and non-clinical controls achieving a mean score of 109.12 (SD = 2.26), $F(1, 48) = 1.27$, $p = .27$. Note that, contrary to many of the previous studies in this thesis, the direction of the mean general ability scores suggests slightly higher ability in the patient group as compared to the non-clinical control group. Correlational analyses indicated that depression, $r(48) = .81$, $p < .001$, anxiety, $r(48) = .83$, $p < .001$, and stress scores,

$r(48) = .82$, $p < .001$, were all highly positively correlated with scores on the OCI-R. High positive correlations between symptoms of OCD, depression, anxiety and stress are not unusual (see Taylor, 1998 for a review); for example, Foa et al. (2002) noted a correlation of .70 between OCI-R scores and scores on the Beck Depression Inventory (Beck & Steer, 1987). These relationships are commonly attributed to a true symptom overlap between symptoms of OCD, depression and anxiety (Gönner, Leonhart, & Ecker, 2008). Further, these correlations are expected of a participant group where participants were included as long as OCD was their primary diagnosis (i.e., participants who met criteria for another anxiety or depressive disorder were not excluded). Given the high levels of association between OCD symptoms and symptoms of depression, anxiety, and stress, controlling for these symptom levels was not possible, as this would remove much of the same error variance from the dependent variable (Stevens, 1996).

Premise Monotonicity Induction-Deduction Task

Preliminary analyses. ANOVA analyses were conducted on the effect of the order in which participants received instructions (deduction or induction first) on the proportion of positive (‘strong’/ ‘valid’) responses. There was no main effect of instruction order, $F(1, 45) = 2.31$, $p = .14$, nor did instruction order interact with reasoning condition, $F(1, 46) = 1.38$, $p = .25$. Therefore, all subsequent analyses were collapsed across instruction order.

ANOVA analyses were also conducted on the effect of conclusion type (general or specific) on the proportion of positive (‘strong’ / ‘valid’) responses to invalid arguments. There was a main effect of conclusion type, $F(1, 48) = 10.66$, $p = .002$, $\eta_p^2 = .18$, such that participants were more likely to endorse invalid arguments when these featured a general ($M = 0.32$, $SD = 0.27$) than specific conclusion ($M = 0.21$, $SD = 0.23$). Conclusion type interacted with argument length, $F(1, 48) = 4.56$, $p = .04$, η_p^2

=.09, such that longer arguments were accepted more often than shorter ones, and this effect was larger for general conclusions. There was also a three-way interaction between reasoning condition, conclusion type and argument length, $F(1, 48) = 12.26$, $p = .001$, $\eta_p^2 = .20$. This was because the larger effect of argument length for general conclusions only applied under induction, $F(1, 48) = 11.66$, $p = .001$, $\eta_p^2 = .20$, but not deduction instructions, $F(1, 48) = 0.19$, $p = .66$.

Collectively, and in line with Feeney (2007), this suggests that the inductive heuristic of premise monotonicity was stronger for arguments that featured a general, rather than a specific, conclusion. Nevertheless, because an effect of premise monotonicity was found for both general, $F(1, 48) = 28.30$, $p < .001$, $\eta_p^2 = .37$, and specific conclusions, $F(1, 48) = 38.04$, $p < .001$, $\eta_p^2 = .44$, all further analyses were collapsed across conclusion type (see Appendix D for full analyses comparing specific and general conclusions).

Proportion of positive (‘strong’/ ‘valid’) responses. These analyses examined the proportion of positive responses for the eight argument types (valid and invalid at each level of argument length and OCD-relevance); that is, the proportion of items that were evaluated as ‘strong’ (induction instructions) or ‘valid’ (deduction instructions) in binary judgments. Estimated marginal means for valid and invalid items are shown in Figure 5.1. The proportion of positive responses was analysed in a 2 (clinical group: OCD patient or non-clinical control) × (2) (reasoning condition: induction or deduction) × (2) (item validity: valid or invalid) × (2) (argument length: one-premise or five-premise) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA with repeated measures on the last four factors and individual general ability scores entered as a covariate (see Appendix D for full analyses).

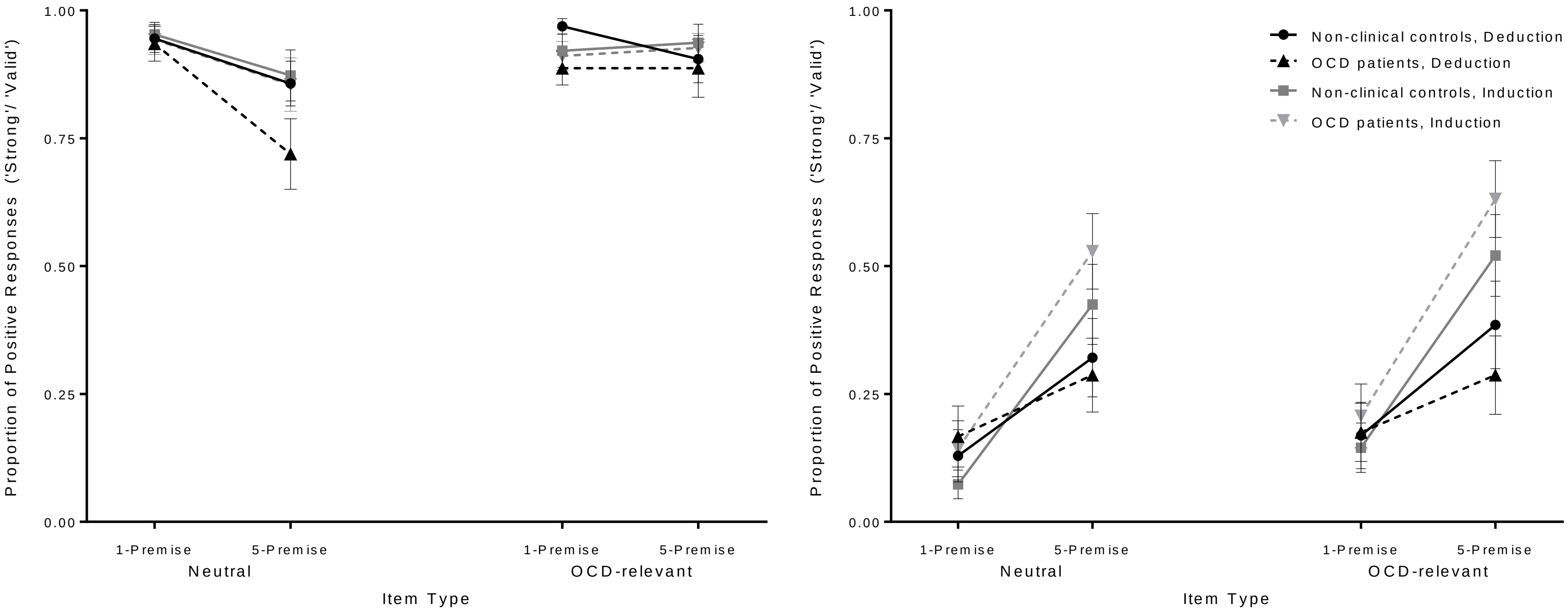

Figure 5.1. Estimated marginal mean proportion of positive responses (‘strong’/ ‘valid’) to valid (left panel) and invalid (right panel) items for two argument length levels by clinical group, OCD-relevance and reasoning condition on the Experiment 8 premise monotonicity induction-deduction task. Error bars represent standard error of the mean.

Analyses revealed a main effect of validity, such that valid items were more likely to be accepted than invalid items, $F(1, 48) = 288.91$, $p < .001$, $\eta_p^2 = .86$. Contrary to Rotello and Heit (2009) but consistent with Experiment 1, the effect of item validity only showed a weak trend toward differing between reasoning conditions, $F(1, 48) = 2.59$, $p = .11$. The pattern of means under the different reasoning instructions suggests that participants trended toward better discrimination between valid and invalid arguments when making deductive ($M = 0.65$, $SD = 0.31$) rather than inductive ($M = 0.58$, $SD = 0.28$) judgments.

Valid items. When analyses were restricted to a consideration of valid items alone, there was no main effect of reasoning condition, $F(1, 48) = 1.38$, $p = .25$, nor OCD-relevance, $F(1, 48) = 2.90$, $p = .10$. This indicates that valid arguments were equally likely to be accepted under deduction and induction instructions, and regardless of whether they were OCD-neutral or OCD-relevant. There was, however, a main effect of clinical group, such that OCD patients were less willing to accept valid arguments than non-clinical controls, $F(1, 47) = 4.36$, $p = .04$, $\eta_p^2 = .09$[11].

As can be seen in the left-hand panel of Figure 5.1, shorter arguments (i.e., with only one premise) were more likely to be accepted than longer arguments, $F(1, 48) = 9.67$, $p = .003$, $\eta_p^2 = .17$. This is consistent with both the findings of Rotello and Heit (2009) and Experiment 1. Argument length also interacted with OCD-relevance, such that the decrease in argument acceptance with increasing argument length was larger for OCD-neutral than for OCD-relevant items, $F(1, 48) = 10.01$, $p = .003$, $\eta_p^2 = .17$. No other interactions reached significance.

[11] This group difference disappeared when general ability was not controlled however, $F(1, 48) = 1.89$, $p = .18$. It is possible that this reflects the slightly higher general ability scores noted in the OCD patient group counteracting any impairment in deductive reasoning.

Invalid items. As can be seen in the right-hand panel of Figure 5.1, increasing argument length increased acceptance of invalid items, $F(1, 48) = 50.25$, $p < .001$, $\eta_p^2 = .51$, and this effect was stronger under induction instructions than deduction instructions, $F(1, 48) = 17.19$, $p < .001$, $\eta_p^2 = .26$. This replicates the premise monotonicity effect noted in previous studies (e.g., Experiment 1, Heit & Rotello, 2012; Rotello & Heit, 2009).

Participants were more likely to accept invalid arguments under induction than deduction instructions, $F(1, 48) = 6.33$, $p = .02$, $\eta_p^2 = .12$. Participants were also more likely to accept invalid items that contained OCD-relevant rather than OCD-neutral content, $F(1, 48) = 10.37$, $p = .002$, $\eta_p^2 = .18$. There was no main effect of clinical group, $F(1, 47) = 0.41$, $p = .53$, indicating that OCD patients and non-clinical controls did not differ in the proportion of invalid items they accepted. No other interaction effects reached significance.

Confidence ratings. Estimated marginal mean confidence ratings for each clinical group and argument type under both deductive and inductive reasoning instructions are shown in Table 5.4.

Valid items. Participants were more confident making judgments about shorter ($M = 4.60$, $SD = 0.50$) than longer ($M = 4.47$, $SD = 0.51$) arguments, $F(1, 48) = 14.32$, $p < .001$, $\eta_p^2 = .23$. This effect was larger for items that were OCD-neutral ($M = 4.56$, $SD = 0.52$) than for OCD-relevant ($M = 4.51$, $SD = 0.51$) items, $F(1, 48) = 6.44$, $p = .014$, $\eta_p^2 = .12$. There was no main effect of reasoning condition, $F(1, 48) = 0.02$, $p = .89$, clinical group, $F(1, 48) = 4.16$, $p = .17$, or OCD-relevance, $F(1, 48) = 1.85$, $p = .18$. No other interactions reached significance. In sum, for valid items, confidence in argument evaluations did not differ as a function of reasoning instruction, clinical status, or the OCD-relevance of items.

Invalid items. For invalid items, participants were also more confident making judgments about shorter (M = 4.19, SD = 0.68) rather than longer (M = 3.92, SD = 0.72) arguments, $F(1, 48) = 19.76$, $p < .001$, $\eta_p^2 = .29$. No other main effects or interactions reached significance.

Table 5.4

Estimated Marginal Mean Confidence Ratings (out of 5) by Clinical Group, Reasoning Condition, OCD-Relevance, Argument Validity and Argument Length on the Experiment 8 Premise Monotonicity Induction-Deduction Task (Standard Deviations in Parentheses)

		Non-clinical controls		OCD patients	
Validity	Length	Deduction	Induction	Deduction	Induction
OCD-neutral items					
Valid	One-premise	4.72 (0.55)	4.76 (0.42)	4.54 (0.70)	4.62 (0.59)
	Five-premise	4.60 (0.61)	4.60 (0.59)	4.26 (0.76)	4.39 (0.60)
Invalid	One-premise	4.34 (0.68)	4.20 (0.72)	4.11 (0.88)	4.16 (0.78)
	Five-premise	4.13 (0.84)	4.01 (0.67)	3.93 (0.90)	3.72 (0.83)
OCD-relevant items					
Valid	One-premise	4.65 (0.58)	4.54 (0.61)	4.50 (0.65)	4.47 (0.53)
	Five-premise	4.56 (0.57)	4.57 (0.59)	4.41 (0.69)	4.37 (0.70)
Invalid	One-premise	4.40 (0.76)	4.09 (0.79)	4.14 (0.86)	4.10 (0.67)
	Five-premise	4.02 (0.92)	4.13 (0.70)	3.79 (0.92)	3.65 (0.72)

Forced-Choice Premise Diversity Task

The key outcome measures were: 1) the proportion of diverse premise pairs endorsed as stronger evidence for a given conclusion (see Table 5.5), and 2) perceived usefulness ratings assigned to the diverse and non-diverse premise pairs (see Table 5.6 and Appendix D for full analyses).

Proportion of diverse arguments endorsed. Choice data were analysed in a 2 (clinical group: OCD patients or non-clinical controls) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA comparison with repeated measures on the last factor, controlling for general ability. As can be seen in Table 5.5, both non-clinical controls and OCD patients were more likely to prefer diverse evidence in support of given conclusions than would be expected by chance, for both OCD-neutral and OCD-relevant items (all *p*'s < .05). Table 5.5 shows that OCD patients tended to endorse diverse arguments less frequently than non-clinical controls, but this effect was not statistically reliable, $F(1, 47) = 0.39$, $p = .53$. Participants were more likely to endorse the diverse arguments when item content was OCD-neutral as opposed to OCD-relevant, $F(1, 48) = 4.97$, $p = .03$, $\eta_p^2 = .09$. However, this effect did not interact with clinical group, $F(1, 48) = 0.18$, $p = .68$. There were no group differences in preference for the diverse arguments across either OCD-neutral or OCD-relevant items.

Table 5.5

Estimated Marginal Mean Proportion of Diverse Arguments Endorsed by Clinical Group and OCD-Relevance on the Experiment 8 Forced-Choice Premise Diversity Task (Standard Deviations in Parentheses) and Details of Significance Tests Compared to Chance Responding (0.50)

	Non-clinical controls				OCD patients			
OCD-relevance	Proportion diverse	*t*	*df*	*p*	Proportion diverse	*t*	*df*	*p*
Neutral	0.77 (0.32)	3.81	24	.001**	0.71 (0.32)	3.52	24	.002**
Relevant	0.71 (0.28)	3.34	24	.003**	0.67 (0.31)	3.05	24	.005**

Note. $^*p < .05$, $^{**}p < .01$.

Ratings of perceived usefulness. These data were analysed in a 2 (clinical group: OCD or non-clinical control) × (2) (item diversity: diverse or non-diverse) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) ANCOVA, with repeated measures on the last two factors. Overall, participants rated diverse arguments as more useful than those non-diverse arguments, $F(1, 48) = 9.26$, $p = .004$, $\eta_p^2 = .16$. Importantly, as can be seen in Figure 5.2, there *were* group differences in the tendency to rate diverse evidence as stronger than non-diverse evidence, $F(1, 48) = 5.47$, $p = .02$, $\eta_p^2 = .10$. Simple effects analyses using a Bonferroni-adjusted alpha level of .0125 demonstrated that the non-clinical control group gave higher usefulness ratings to diverse arguments over non-diverse arguments, $F(1, 48) = 14.48$, $p < .001$, $\eta_p^2 = .23$, but OCD patients did not, $F(1, 48) = 0.25$, $p = .62$. Only the non-clinical control group showed a robust preference for diverse evidence. This replicates the main result of Experiment 3, albeit with a ratings measure rather than a choice measure. There were no clinical group differences in usefulness ratings assigned to either diverse, $F(1, 47) = 2.32$, $p = .14$, or non-diverse arguments, $F(1, 47) = 2.38$, $p = .13$. There was also a main effect of OCD-relevance, such that arguments relating to OCD-relevant content were generally rated as more useful than arguments relating to OCD-neutral content, $F(1, 48) = 9.48$, $p = .003$, $\eta_p^2 = .17$. No other main effects or interactions reached significance.

Table 5.6

Estimated Marginal Mean Perceived Usefulness Ratings (out of 10) by Clinical-Group and OCD-Relevance on the Experiment 8 Forced-Choice Premise Diversity Task (Standard Deviations in Parentheses)

OCD-relevance	Non-clinical controls		OCD patients	
	Diverse	Non-diverse	Diverse	Non-diverse
Neutral	6.29 (1.83)	4.82 (1.68)	5.65 (1.73)	5.46 (1.53)
Relevant	6.33 (1.71)	5.16 (1.59)	6.08 (1.63)	5.94 (1.49)

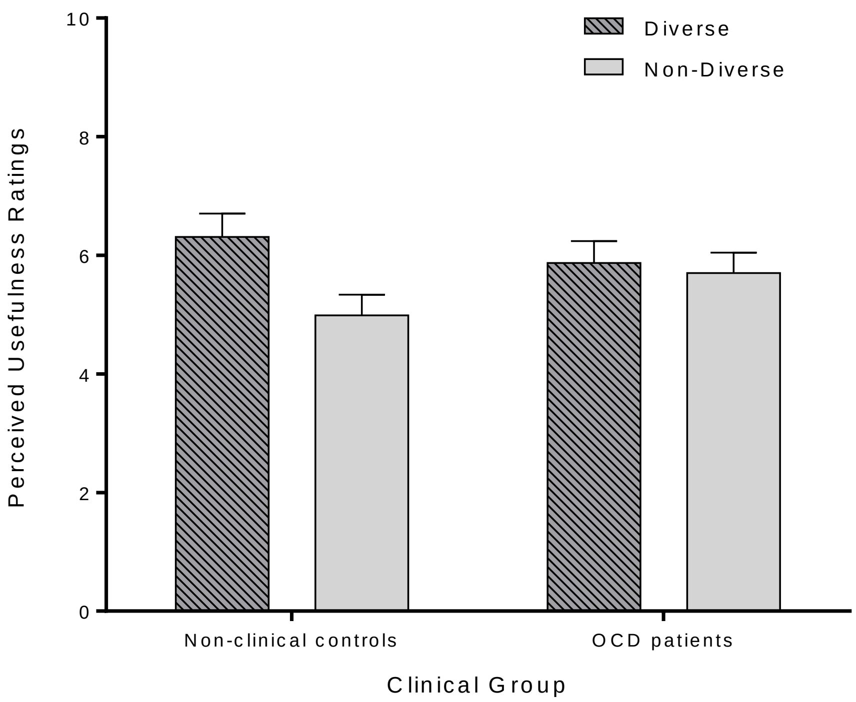

Figure 5.2. Estimated marginal mean perceived usefulness ratings (out of 10) for diverse and non-diverse arguments by clinical group on the Experiment 8 forced-choice premise diversity task. Error bars represent standard error of the mean.

Similarity Ratings Task

Proportion of previously nominated ‘non-diverse’ items endorsed as similar. As can be seen in Table 5.7, both non-clinical controls and OCD patients were more likely to endorse the ‘non-diverse’ premises as being the most similar premise pair than would be expected by chance; this was true of both OCD-neutral and OCD-relevant items (all *p*’s < .001; see Appendix D for full analyses).

A 2 (clinical group: OCD patient or non-clinical control) × (2) (OCD-relevance: OCD-neutral or OCD-relevant) repeated measures ANCOVA, including general ability as a covariate, was conducted on the proportion of previously nominated ‘non-diverse’ items that were endorsed as being the most similar option. Henceforth, these responses are referred to as ‘matching’ responses.

Overall, participants were more likely to offer matching responses when item content was OCD-neutral rather than OCD-relevant, $F(1, 48) = 24.26$, $p < .001$, $\eta_p^2 = .34$. However, the proportion of matching responses did not differ as a function of clinical status, $F(1, 47) = 0.61$, $p = .44$. No interaction effects reached significance.

Table 5.7

Estimated Marginal Mean Proportion of ‘Matching’ Responses (Nominally ‘Non-Diverse’ Items Rated as Similar) by Clinical Group and OCD-Relevance on the Experiment 8 Similarity Ratings Task (Standard Deviations in Parentheses) and Details of Significance Tests Compared to Chance Responding (0.50)

	Non-clinical controls				OCD patients			
OCD-relevance	Proportion 'matching'	*t*	*df*	*p*	Proportion 'matching'	*t*	*df*	*p*
Neutral	0.97 (0.05)	45.96	24	<.001**	0.99 (0.03)	73.00	24	<.001**
Relevant	0.92 (0.07)	29.20	24	<.001**	0.94 (0.06)	34.49	24	<.001**

Note. **p* < .05, ***p* < .01.

Similarity ratings. A 2 (clinical group: OCD patient or non-clinical control) × (2) ('diversity' status: previously nominated 'diverse' or 'non-diverse') × (2) (OCD-relevance: OCD-neutral or OCD-relevant) repeated measures analysis controlling for general ability was also conducted on the similarity ratings given to OCD-relevant and OCD-neutral previously nominated 'non-diverse' and 'diverse' premises. Estimated marginal means can be seen in Table 5.8.

As can be seen, previously nominated 'non-diverse' premises were rated as more similar than previously nominated 'diverse' premises, $F(1, 48) = 217.34$, $p < .001$, $\eta_p^2 = .82$. Overall, OCD-relevant premise pairs were given higher similarity ratings than OCD-neutral premise pairs, $F(1, 48) = 28.53$, $p < .001$, $\eta_p^2 = .37$. OCD-relevance also interacted with the nominated diversity of the items, $F(1, 48) = 17.19$, $p < .001$, $\eta_p^2 = .26$. Table 5.8 shows that there was a larger difference between similarity ratings for 'diverse' and 'non-diverse' items when item content was OCD-neutral. There was no main effect of clinical group on overall ratings of similarity, $F(1, 47) = 0.00$, $p = .95$, nor did clinical status interact with the previously nominated diversity of the premises, $F(1, 48) = 0.07$, $p = .79$.

Table 5.8

Estimated Marginal Mean Similarity Ratings (out of 10) Given to Previously Nominated 'Diverse' and 'Non-Diverse' Premises by Clinical Group and OCD-Relevance on the Experiment 8 Similarity Ratings Task (Standard Deviations in Parentheses)

	Non-clinical controls		OCD patients	
OCD-relevance	'Diverse'	'Non-diverse'	'Diverse'	'Non-diverse'
Neutral	3.67 (1.99)	7.76 (1.15)	3.74 (1.99)	7.71 (1.23)
Relevant	4.41 (1.57)	8.00 (0.93)	4.55 (1.55)	7.95 (0.98)

Discussion

This experiment examined whether the deductive and inductive deficits found in analogue samples of individuals high in OCD symptomology would also be found in a clinically diagnosed sample. The results for each of the various reasoning tasks are discussed in turn.

Premise Monotonicity

OCD patients accepted fewer valid arguments than did non-clinical controls on the premise monotonicity task when general ability was controlled, but were no different from controls in their endorsement of invalid arguments. This is a partial replication of the reduced sensitivity to validity found in the high OCD analogue sample in Experiment 1. Note that sensitivity in distinguishing between valid and invalid arguments (proportion of valid arguments accepted – proportion of invalid arguments accepted) was higher in Experiment 1 (*M* = 0.70, *SD* = 0.26) than in the current experiment (*M* = 0.61, *SD* = 0.26). This is likely to be related to the within-subject manipulation of reasoning condition in the current experiment. Even though participants evaluated items deductively or inductively in separate blocks, the need to switch between the two criteria may have reduced sensitivity to the differences between valid and invalid items. This made it more difficult to detect clinical group differences in sensitivity to validity in the current experiment.

As expected, OCD patients were as sensitive to argument length in their judgments of inductive plausibility as non-clinical controls. That is, both groups showed a reliable premise monotonicity effect under induction instructions. OCD patients and non-clinical controls were also equally confident in their argument evaluations, and OCD patients were not more confident when reasoning deductively than inductively.

Premise Diversity

Both OCD patients and non-clinical controls saw diverse premises as promoting stronger inductive generalisation in forced-choice decisions. Notably however, non-clinical controls (but not OCD patients) rated diverse premise pairs as stronger than non-diverse pairs. This replicates a major finding of Experiment 3, where ratings of perceived usefulness given by individuals high in OCD symptoms did not differ for non-diverse and diverse premise pairs. This is also a conceptual replication of the diversity effect demonstrated in Experiments 2 and 3, where those highest in OCD symptoms exhibited a reduced preference for diverse evidence, albeit on a forced-choice measure.

It is not entirely clear why this experiment found group differences in preference for diverse evidence as indexed by evidence ratings, but not when indexed by forced-choice decisions. It should be noted, however, that in previous studies of inductive reasoning with non-clinical samples, ratings of argument strength and forced-choice responses do not always align. For example, Osherson et al. (1990) found evidence of premise-conclusion asymmetry in forced-choice responses but not in argument strength evaluations. Relatedly, the current experiment generally found a weaker correspondence between individuals' endorsements of diverse premise pairs based on forced-choice and ratings measures than was found in Experiment 3 (Experiment 3, $r(98) = .38$; Experiment 8, $r(48) = .29$).

Similarity Ratings

Finally, and contrary to the results of Experiment 5, no differences were noted between OCD patients and non-clinical controls on perception of similarity between premise pairs, in either forced-choice selections or similarity ratings. One caveat is that in the current experiment, similarity ratings were obtained *after* the premise diversity task had been completed. As the premises used in both tasks were identical, it is possible that diversity judgments may have influenced judgments of similarity (e.g., processing the diversity of an item may have primed deeper processing of similarity).

Notwithstanding the procedural differences between studies, the current results suggest that differences in sensitivity to diversity between OCD patients and non-clinical controls are not driven by differences in perceptions of premise similarity. A related disjunct between diversity and similarity was found in the earlier experiments. In Experiment 5, differences in the perception of similarity shown by the group high in OCD symptomatology were much larger for OCD-relevant than for OCD-neutral items. This was not reflected in the diversity results (Experiment 3), where the reduced preference for diversity exhibited by individuals high in OCD symptomatology held equally across *both* kinds of items. Other possible mechanisms that may underlie clinical group differences in sensitivity to evidence diversity are considered in the General Discussion (Chapter 6).

Summary

Experiment 8 showed that a clinically diagnosed group of OCD patients were less sensitive to the implications of premise diversity (as indexed by perceived usefulness ratings assigned to diverse and non-diverse premises) than were non-clinical controls. This reduced sensitivity to premise diversity is unlikely to be explained by group differences in the perception of similarity. Further, the reduced sensitivity of those high in OCD symptomatology to argument validity was partially replicated.

CHAPTER 6 - General Discussion

Summary of Thesis Aims

The overall aim of this thesis was to advance current understanding of reasoning deficits in OCD, particularly in the areas of deductive and inductive reasoning. Previous research has suggested that OCD patients seem to be impaired, relative to non-clinical and anxious controls, on tasks of inductive reasoning (Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007). On tasks of deductive reasoning, those with OCD have generally performed similarly to non-clinical and anxious controls (Pélissier & O'Connor, 2002), with one notable exception (Simpson et al., 2007).

A major limitation of this previous work is that comparisons of deductive and inductive reasoning were conducted on reasoning tasks that differed in a number of other respects, such as task difficulty, item content, and whether other reasoning processes (outside of deduction or induction) were necessary for correct performance. Furthermore, the inductive tasks used were generally not well-established in the wider reasoning literature, making it harder to draw inferences about the reasoning processes that underlie performance.

Thus, one of the aims of the current thesis was to investigate deductive and inductive reasoning in OCD on tasks that allowed for a direct comparison between the two reasoning processes. This thesis utilised a well-established procedure to facilitate this comparison (Heit & Rotello, 2005; Heit et al., 2012; Rips, 2001; Rotello & Heit, 2009), which involved presenting participants with a common stimulus set and either deductive or inductive instructions.

An additional aim was to conduct a broader examination of hypothesised reasoning deficits in OCD. Recall that there are at least two distinct factors that influence the way in which one makes inductive inferences (see Hayes et al., 2010 for a review): inductive sampling heuristics (or general rules for deciding whether information is generalisable) and background knowledge. This thesis examined how the performance of individuals with differing levels of OCD symptomatology was affected by each of these factors, with the inductive sampling heuristics of premise monotonicity (sample size) and premise diversity, and the influence of background causal knowledge on inductive judgments being the main topics of investigation. Deductive validity was also manipulated in multiple ways, using both class-inclusion and other argument structures (i.e., modus ponens, conjunctive syllogisms, and categorical syllogisms).

The majority of experiments (Experiments 2-5, 7-8) included both OCD-neutral and OCD-relevant items, in order to determine whether any deficits found would be global across all item types, or specific to one particular type of item. In addition, general ability was measured and controlled in Experiments 2-8. The issues outlined above were examined with two types of OCD samples: analogue populations (Experiments 1-7) and a clinically diagnosed sample (Experiment 8).

Summary of the Major Empirical Findings

A summary of reasoning performance across the eight experiments in those high in OCD symptoms (as defined by clinical cut-offs on the OCI-R, OCI-R quartiles, or clinical diagnosis) is shown in Table 6.1.

Table 6.1

Summary of Obtained Deductive and Inductive Performance of Those High in OCD Symptoms, as Defined by Clinical Cut-off (≥21 on the OCI-R), the Highest Quartile of OCD Symptoms, or Clinical Diagnosis

			Definition of high OCD sample		
Reasoning	Task	Experiments	OCI-R cut-off	OCI-R top quartile	Clinical diagnosis
Deduction	Class-inclusion	1, 2, 8	X	X	X
	Modus ponens/conjunctive syllogisms	6	✓	✓	N/A
	Categorical syllogisms	7	✓	✓	N/A
Induction	Monotonicity	1, 8	✓	✓	✓
	Diversity	2, 3, 8	?	X	X
	Background knowledge	6	✓	✓	N/A

Note. ✓ = intact performance; X = impaired performance; ? = some evidence for impairments; N/A = task was not administered. In Experiments 2-8, any group differences noted existed above group differences in general ability. All group differences were global across both OCD-neutral and OCD-relevant content.

Deductive Reasoning in Obsessive-Compulsive Disorder

The deductive performance of individuals high in OCD symptoms relative to individuals low in OCD symptoms was mixed, and depended on the way that deductive validity was manipulated. Impairments were noted on judgments of deductive validity, but only when validity was manipulated via class-inclusion (i.e., where arguments were only valid if a superordinate category was included as one of the premises). This was first noted in Experiment 1, in which only OCD-neutral content was employed. In this experiment, individuals high in OCD symptoms, as defined by both the clinical cut-off (Foa et al., 2002) and by the upper quartile, were less able to distinguish between valid and invalid arguments relative to those low in OCD symptoms. That is, these individuals were *less* likely to accept valid arguments and *more* likely to accept invalid arguments relative to those low in symptoms. Obsessive-compulsive symptoms were a better predictor of sensitivity to argument validity than was the extent to which individuals were intolerant of uncertainty.

This relationship between OCD and impaired sensitivity to argument validity via class-inclusion was replicated in Experiment 2, where those highest in OCD symptoms were less able to distinguish between valid and invalid arguments compared to those lowest in OCD symptoms. This deficit appeared across both OCD-neutral and OCD-relevant materials, and importantly, was found even after controlling for general ability. Both OCD groups were more sensitive to argument validity under deductive rather than inductive instructions. Finally, in Experiment 8, OCD patients accepted fewer valid class-inclusion arguments than did a non-clinical control group. Again, this was a global effect found across both OCD-neutral and OCD-relevant content. Note however, that although those high in OCD symptoms consistently were poorer at distinguishing between valid and invalid class-inclusion arguments across multiple studies, these individuals were still more likely overall to accept valid over invalid arguments.

When deductive validity was manipulated via logical structures other than class-inclusion (i.e., modus ponens, conjunctive syllogisms and categorical syllogisms), no differences were found between those high and low in OCD symptoms in sensitivity to argument validity. On the causal consistency induction-deduction task, which included arguments based on modus ponens and conjunctive syllogisms (Experiment 6), participants were more sensitive to the validity of an argument when given deductive rather than inductive instructions. However, in this case, individuals high in OCD symptoms were just as able to distinguish between valid and invalid arguments as those low in symptoms.

This pattern of equivalent group performance was also found on a belief bias task using categorical syllogisms (Experiment 7), with both high and low OCD groups equivalent in their ability to correctly follow arguments to a logical conclusion, or to indicate when premises did not necessitate a particular conclusion. Belief bias (as defined by a belief by validity interaction) was only noted on OCD-neutral items, and did not vary as a function of OCD group. Given that no belief by validity interaction was noted for OCD-relevant items, it is unclear whether belief bias plays a role in preventing the disconfirmation of the anxious beliefs held by those high in OCD symptoms, as previously hypothesised by de Jong et al. (1997) and Vroling and de Jong (2009).

Possible explanations regarding why group differences in deductive reasoning were found on some tasks but not others are examined in the Theoretical Implications section below.

Inductive Reasoning in Obsessive-Compulsive Disorder

Relative to individuals low in OCD symptoms, the inductive abilities of those high in OCD symptoms varied depending on whether such reasoning was based on inductive sampling heuristics or background knowledge. In terms of inductive sampling heuristics, those high and low in OCD symptoms were well able to use premise monotonicity (sample size) as a guide for their inductive judgments. Sensitivity to sample size was found in both analogue (Experiment 1) and clinical (Experiment 8) OCD groups, and across both OCD-neutral and OCD-relevant content. This preference for inductive generalisation based on greater amounts of positive evidence is consistent with the performance of non-clinical samples in previous studies of inductive reasoning (e.g., Feeney, 2007; Heit & Rotello, 2012; Osherson et al., 1990; Rotello & Heit, 2009). Further, in line with Heit and Rotello (2012) and Rotello and Heit (2009), the sample size effect was stronger under induction than deduction instructions.

An important finding of the thesis, however, was that individuals high in OCD symptomatology appeared impaired in their use of evidence diversity. More diverse evidence is usually seen as a stronger basis for inductive generalisation than less diverse evidence, when premises and conclusions belong to the same superordinate category (e.g., Kim & Keil, 2003; Kincannon & Spellman, 2003; López, 1995; Osherson et al., 1990). Relative to individuals low in OCD symptoms, however, those high in OCD symptoms exhibited a reduced preference for diverse evidence when making their inductive inferences across OCD-neutral and OCD-relevant content (although these individuals did still endorse diverse evidence above chance levels). This reduced sensitivity to diversity was revealed across two main measures: reduced selection of diverse premise pairs in forced-choice decisions (Experiment 2 and 3) and in ratings of perceived usefulness of the premise pairs (Experiments 3 and 8). This pattern was observed in both analogue samples and in a clinically diagnosed sample.

This represents the first empirical evidence of a systematic difference between those high and low in OCD symptoms on a well-established inductive heuristic. Importantly, this reduced sensitivity to premise diversity appeared to be a global deficit occurring across both OCD-neutral and OCD-relevant items. Moreover, the effect was evident over and above group differences in general ability. One caveat is that this reduced sensitivity to premise diversity did not appear when participants high in OCD symptoms were *searching* for further evidence to support an inductive inference, rather than just evaluating arguments (Experiment 4). This indicates that, although individuals high in OCD symptomatology may be less willing to generalise properties on the basis of diverse premises than are individuals low in OCD symptomatology, they may be just as likely to utilise diversity when testing a hypothesis. However, as was discussed in Chapter 3, the truth of the to-be-tested premises in the hypothesis testing task was open to interpretation, which may have obscured group differences in the use of diversity.

One factor that may contribute to OCD group differences in sensitivity to evidence diversity may be differences in perception of the relative similarity of argument premises. In Experiment 5, individuals high in OCD symptoms rated the similarity between premises used to generate the diverse and non-diverse item sets in a way that differed from those low in OCD symptoms. In other words, the inductive items that were seen as nominally ‘diverse’ by those low in OCD symptoms may not have been perceived as such by those high in OCD symptoms.

However, although group differences in the perception of similarity were larger for OCD-relevant than for OCD-neutral items, the reduction in sensitivity to diversity exhibited by those high in OCD symptoms was *not* larger for OCD-relevant items.

Additionally, although evidence of a reduced use of the diversity heuristic was found in the clinical OCD sample (Experiment 8), differences in the perceived similarity of premises were not replicated in this sample. Collectively, this implies that a reduced sensitivity to diversity cannot be completely explained by OCD group differences in the perception of similarity. Rather, it seems likely that group differences in sensitivity to diversity reflects differences in the value that the high and low OCD symptom groups assign to diverse and non-diverse evidence. Possible explanations for the reduced sensitivity to diversity are further discussed in the Theoretical Implications section.

Finally, it was found that individuals high in OCD symptomatology were just as able to use causal knowledge when making judgments of inductive plausibility as were individuals low in OCD symptomatology (Experiment 6). Further, and in line with Rips (2001), both high and low OCD groups used causal consistency as a basis for reasoning to a greater extent when making inductive as opposed to deductive judgments.

The Role of General Ability in Inductive and Deductive Reasoning in Obsessive-Compulsive Disorder

General ability was measured in the majority of experiments in this thesis, primarily because higher levels of general ability are related to improved deductive performance and greater use of inductive sampling heuristics (e.g., Feeney, 2007; Stanovich & West, 1998a). This was true in the current series of experiments. For example, sensitivity to argument validity in Experiment 2 (as defined by the proportion of valid arguments endorsed - the proportion of invalid arguments endorsed) was moderately positively correlated with general ability scores, $r(114) = .34$. Similarly, the tendency to endorse diverse premises as being stronger evidence for a particular conclusion in Experiment 3 was positively correlated with general ability, $r(98) = .21$.

Previous studies of general ability in OCD have often indicated that patients do not differ from non-clinical controls, or are only impaired on tasks of non-verbal reasoning (e.g., Bannon et al., 2002; Bédard et al., 2009; Boone et al., 1991; Bucci et al., 2007; Nakao et al., 2009; Veale et al., 1996). In the current thesis, patterns of group performance on the WASI were at times consistent with previous research, with no differences in general ability noted between analogue participants high or low in OCD symptoms (Experiment 2), or between OCD patients and non-clinical controls (Experiment 8). However, those high in OCD symptoms did at times exhibit lower full-scale general ability estimates than those low in OCD symptoms (the sample used in Experiments 3 and 6, and the sample used in Experiments 4 and 7). Further, lower general ability scores were noted for those high in OCD symptoms on a proxy measure of general ability, relative to those low in OCD symptoms (Experiment 5).

Although general ability was usually correlated with reasoning performance, the obtained group differences in reasoning (e.g., in sensitivity to diversity) could not be reduced to group differences in general ability. In the majority of cases, reasoning differences in individuals high and low in OCD symptoms persisted when the effects of general ability were controlled in analyses of covariance. This is an important novel finding: apart from one study which matched their non-clinical control and OCD patient group on premorbid general ability and years educated (Simpson et al., 2007), no other study investigating induction and deduction in OCD has previously accounted for the possible role of general ability.

There were two relatively minor exceptions to this trend. First, when selecting further evidence to test a conclusion in Experiment 4, individuals high in OCD symptoms selected fewer diverse options and more non-diverse options relative to those low in symptoms, but only when general ability was not controlled. Second, in

Experiment 8, OCD patients accepted fewer valid items relative to non-clinical controls on the premise monotonicity task, but only when general ability was controlled. The latter finding may have been explained by the slightly higher general ability scores of OCD patients offsetting any impairment in deductive reasoning.

Overall, these findings highlight the importance of including measures of general ability in comparisons of reasoning performance between groups that vary in clinical symptomology. Where possible, future work should build on the existing findings by employing more elaborate ability assessment instruments (e.g., Wechsler Adult Intelligence Scale-IV; Wechsler, 2008) that allow for separate reliable estimates of verbal and performance abilities.

Theoretical Implications

This section examines the theoretical implications of the current findings for the "spared deduction, impaired induction" account of OCD. The implications of the global or specific nature of these deficits are also discussed.

Induction and Deduction in Obsessive-Compulsive Disorder: the "Spared Deduction, Impaired Induction" Account

The Inference-Based Approach (IBA) model of OCD (O'Connor et al., 2005) proposes that obsessions are actually inferences derived through faulty inductive reasoning. Further, the model suggests that deductive reasoning is intact, and that faulty inductive reasoning interacts with intact deductive reasoning in 'obsessional narratives' that convince patients to distrust their senses. That is, this account suggests that in OCD, impairments should be noted in inductive but not deductive reasoning.

The results of the current experiments challenge both components of this "spared deduction, impaired induction" account. Across the eight experiments, individuals high in OCD symptoms (or those receiving a clinical diagnosis) were able to reason inductively in some contexts, and were at other times impaired when reasoning deductively.

Is there evidence for "spared" deduction in obsessive-compulsive disorder? High OCD symptom individuals exhibited a reduced sensitivity to argument validity, but only when this was manipulated using particular kinds of logical structures (i.e., class-inclusion). Why might those high in OCD symptoms be deductively impaired on class-inclusion arguments and not on other manipulations of logical structure (i.e., modus ponens, conjunctive syllogisms, and categorical syllogisms)? It is likely that an explanation lies in the nature of the tasks used. There are at least two clear ways in which manipulations of validity via class-inclusion differ from the manipulations of logical structure used in other experiments. First, the class-inclusion problems are less complex than many of the other syllogisms used. Second, unlike the other logical forms, class-inclusion arguments require representation of hierarchical information (Greene, 1989).

There seems little doubt that solution of class-inclusion problems requires less working memory load than modus ponens arguments, conjunctive syllogisms (Experiment 6), and the all/some/none categorical syllogisms (Experiment 7). Class-inclusion arguments have been shown to require the processing of fewer relations in working memory than many other logical forms (cf. Halford, Andrews, & Jensen, 2002; Zielinski, Goodwin, & Halford, 2010). For low OCD symptom individuals, this difference in complexity was reflected in patterns of sensitivity to validity (i.e., proportion of valid arguments accepted – proportion of invalid arguments accepted); sensitivity was substantially higher on class-inclusion problems (M = 0.68, SD = 0.34) than modus ponens and conjunctive syllogisms (M = 0.41, SD = 0.27), or categorical syllogisms (M = 0.13, SD = 0.20).

Given that class-inclusion arguments are less complex than the other syllogisms used, it seems puzzling that they yielded differences in performance between high and low OCD symptom participants. One explanation is that the *perception* that class-inclusion is a relatively simple form of reasoning may have negatively impacted the deductive performance of individuals high in OCD symptomatology. Such perceptions can lead to subjective feelings of “fluency” during processing (Alter & Oppenheimer, 2009; Oppenheimer, 2008). Under some circumstances, this leads participants to make inferences that involve shallow or incomplete processing of the available information. In contrast, feelings of disfluency can prompt participants to adopt more systematic processing strategies (Alter, Oppenheimer, Epley, & Eyre, 2007). In support of this view, it has been shown that deductive accuracy on both logical syllogisms and class-inclusion arguments increases when problems are presented in a font that is difficult to read (dysfluent condition) rather than a more standard fluent font (Alter et al., 2007; Rotello & Heit, 2009), despite disfluent fonts being more demanding of cognitive resources.

It is therefore possible that the disfluency likely experienced on the more difficult modus ponens, conjunctive syllogisms, and categorical syllogisms (Experiments 6 and 7) promoted the use of systematic, effortful processing strategies. This in turn may have compensated for deficits in the deductive reasoning abilities of individuals high in OCD symptoms only evident on simple deductive tasks (e.g., class-inclusion tasks, which involve more fluent processing). This argument implies that increasing the disfluency of class-inclusion problems should improve the deductive performance of individuals high in OCD symptoms. These individuals should show deductive performance that is more similar to those low in OCD symptoms when class-inclusion problems are presented under dysfluent processing conditions (e.g., a font that is difficult to read).

A second explanation that might account for why individuals high in OCD symptoms are impaired on only certain deductive tasks is that OCD patients often struggle to impose external structure on tasks. For example, OCD patients (Rankins, Bradshaw, & Georgiou-Karistianis, 2005) and individuals high in obsessive-compulsive personality traits (Yovel, Revelle, & Mineka, 2005) have been shown to have difficulty processing hierarchical information on the global-local hierarchical-letters paradigm (Navon, 1977). These individuals are distracted by local aspects of hierarchical stimuli (i.e., smaller letters that comprise a larger letter) that should be ignored when trying to identify more global aspects (i.e., the larger letter). In the current experiments, individuals high in OCD symptomatology may have focused on local aspects of arguments (e.g., specific examples of mammals, fruit, flowers and alcohol), rather than the global overarching category (e.g., 'all mammals have Property X', 'all fruit has Property Y'). This may have interfered with judgments of class-inclusion validity.

Furthermore, OCD patients tend to over-structure their experience, "involving close attention to the details of input, over-specificity, search for information and the deferral of a decision as to completion" (Reed, 1997, p. 184). For example, when sorting OCD-neutral and OCD-relevant decks of cards into groups (e.g., things that go together by 'weight' or 'seriousness of mistakes'), OCD patients sort cards into a greater number of groups and take longer to do so, relative to clinical controls (Pearsons & Foa, 1984). Pearsons and Foa suggested that this response pattern indicated that OCD patients were over-categorising decks by making finer grained distinctions between items. A tendency to over-categorise may have interfered with class-inclusion judgments, with patients rejecting the supplied superordinate categories in favour of more distinctive categories.

Is there evidence for "impaired" induction in obsessive-compulsive disorder? The inductive performance of individuals high in OCD symptoms was mixed. Relative to individuals low in OCD symptoms, those high in symptoms were as able to utilise sample size and background causal knowledge to make judgments of inductive plausibility. However, individuals high in OCD symptoms were less willing than those low in OCD symptoms to generalise on the basis of diverse evidence. Differences in the way that individuals high in OCD symptoms perceive similarity (e.g., Experiment 5) may account for some, but not all (e.g., Experiment 8) of these differences in inductive reasoning. Further, given that a reduced sensitivity to diversity was noted globally across *both* OCD-neutral and OCD-relevant content, it seems unlikely that this impairment was derived by those high in OCD symptoms inferring more specific or idiosyncratic relations that weakened the perceived evidential usefulness of the diverse pairs.

One possible explanation for the reduced sensitivity to evidence diversity noted in individuals high in OCD symptoms is the notion that generalisation according to evidence diversity involves computing how well the premises 'cover' the superordinate category (Osherson et al., 1990). Osherson and colleagues argue that when presented with premises, participants first determine the most specific category that includes the premises and then determine the degree to which the sample covers the category. Thus, the same difficulty that may account for a reduced sensitivity to class-inclusion validity (i.e., a tendency to over-categorise into more distinctive categories; Pearsons & Foa, 1984) may also account for a reduced sensitivity to diversity. If those high in OCD symptomatology are constructing narrower categories that better include non-diverse, rather than diverse premises, then these individuals may be less likely to generalise on the basis of diverse evidence.

A potential difficulty with this argument is that Osherson et al. (1990) claim that utilising premise monotonicity also requires computing coverage, but as was shown in Experiments 1 and 8, individuals high in OCD symptomatology are unimpaired in their use of sample size information. Against this point, it should be noted that more recent accounts of inductive reasoning (e.g., Feeney, 2007; Heit, 2000; Medin et al., 2003) have questioned whether the unitary process of ‘coverage’ provides an adequate explanation of inductive phenomena such as diversity and monotonicity. Medin et al.’s (2003) Relevance Theory, for example, emphasises the importance of identifying the relevant relations linking premise and conclusion categories in induction. The dissociation between monotonicity (both high and low OCD symptom groups are equally sensitive) and diversity performance (high OCD symptom groups less sensitive than low OCD symptom groups) may reflect the differential difficulty of identifying relations involving the amount of evidence, as opposed to the implications of sampling from a broad range of instances, relative to a narrower range.

It is also notable that some studies have found dissociations between the development of monotonicity and diversity effects. For example, Lo, Sides, Rozelle, and Osherson (2002) found that pre-schoolers were sensitive to premise monotonicity in a property induction task, but showed weak or no sensitivity to diversity. This provides further evidence against the view that identical processes drive sensitivity to monotonicity and diversity.

Clearly, the notion that individuals high in OCD symptoms can utilise certain inductive sampling heuristics and causal background knowledge indicates that their inductive abilities are not universally impaired. Based on extensive work on inductive reasoning with non-clinical populations, this thesis has argued that inductive performance involves a number of different processes, including category-based sampling heuristics and background causal knowledge (e.g., see Bright & Feeney, 2014; Hayes et al., 2010). The current results show that those high in OCD symptoms can show impairments in *some forms of induction* (e.g., use of the diversity heuristic), but are relatively unimpaired in others (e.g., using background causal knowledge). Such dissociations highlight the importance for future clinical research of using a variety of inductive tasks that tap into the multiple processes that drive such reasoning.

The relationship between induction and deduction: Single or dual-processes? Some aspects of the current findings are consistent with the notion that different processes underlie inductive and deductive reasoning. In a number of experiments, it was found that certain task factors had differential effects on inductive and deductive performance. For example, increasing sample size increased argument acceptance to a greater extent under induction instructions relative to deduction instructions (Experiments 1 and 8), replicating Heit and Rotello (2012) and Rotello and Heit (2009). Further, increasing argument acceptance with increasing diversity of premises was only noted under induction instructions, not deduction instructions (Experiment 2). This specificity of the influence of premise diversity to induction instructions has never previously been demonstrated. Going in the other direction, those given deduction instructions were more sensitive to argument validity than those given induction instructions; that is, relative to those in the induction condition, individuals given deduction instructions were more likely to endorse valid class-inclusion arguments and less likely to endorse invalid arguments (Experiment 2).

Although consistent with dual-process accounts of reasoning, it has been shown that such ‘single dissociations’ between induction and deduction (i.e., where a factor affects one type of reasoning more than the other) can be explained by a computational model that incorporates a single reasoning process (see Chapter 1; Lassiter & Goodman, 2015). Single-process accounts (including Lassiter & Goodman's 2015 account) have more difficulty, however, explaining the findings of Rips' (2001) study, which were replicated in Experiment 6. In these studies, logical validity was the dominant factor determining responses in deduction, with consistency with causal knowledge playing a secondary role. These roles were reversed in induction. Such findings represent a 'reversed association' (Dunn & Kirsner, 1988) and represent much stronger evidence against a single-process account. The current replication is important evidence in support of this approach, as there exists only one other published study by Singmann and Klauer (2011) demonstrating a similar pattern.

Overall then, the data seem more consistent with a dual-process approach. However, as outlined above, the results suggest an even more complex picture, with a variety of processes involved in deductive and inductive reasoning. Not only was evidence found of multiple forms of induction, with only some forms impaired in OCD, but differential impairments in deduction were also found, dependant on how arguments were logically structured.

Are Deductive and Inductive Differences Global or Specific to OCD-Relevant Content?

Where group differences in deductive and inductive reasoning were noted in the current thesis, individuals high in OCD symptoms were *globally* impaired. That is, the reduced sensitivity to validity via class-inclusion arguments (Experiments 1, 2 and 8), reduced property generalisation on the basis of premise diversity (Experiment 3), and differences in the perception of premise similarity (Experiment 5) were found across *both* OCD-neutral and OCD-relevant material. This is consistent with previous studies of induction in OCD, which have collectively indicated a global content-independent deficit (Pélissier & O'Connor, 2002; Pélissier et al., 2009; Simpson et al., 2007), with only one exception (Aardema et al., 2009).

Reduced performance across both OCD-neutral and OCD-relevant content implies that the reasoning mechanism underlying performance is impaired, rather than some disorder specific factor (e.g., specific pathological beliefs or anxiety in response to threat-relevant information) influencing performance. Further, this implies that much of the information that the individual reasons about is affected by this deficit, not only content relevant to the individual's obsessional concerns (Taylor, 2002).

Indeed, that these reasoning deficits are domain general may indicate some premorbid factor that causes vulnerability to pathological anxiety (de Jong et al., 1997). Note, however, that it is not possible with the current experimental design to determine whether these reasoning difficulties precede or follow obsessive-compulsiveness. A longitudinal design could distinguish whether deductive and inductive deficits precede the onset of any OCD symptoms. However, the obvious next step in resolving this issue is to establish whether these deficits are specific to OCD or whether they are transdiagnostic (see Limitations of Current Experiments and Directions for Future Research section).

Clinical Implications

This section discusses implications of the obtained findings for the aetiology, maintenance and treatment of OCD. The section begins with consideration of the implications of specific impairments or intact patterns of reasoning noted in the current thesis, before a discussion of suggested modifications to existing cognitive therapies.

Reduced Sensitivity to Premise Diversity

Foa and Kozak (1985) suggest that OCD patients often fail to make inductive generalisations about safety when presented with specific information about the absence of danger; this may exacerbate an individual's experience of OCD symptoms. One mechanism through which reduced safety generalisations may occur is via a reduced sensitivity to premise diversity. For example, if an individual with OCD learns that it is safe to eat dissimilar types of food (e.g., bread and fruit) without wearing food safety gloves, this learning experience might not readily generalise to all hand-held foods, which may interfere with the extinction of already established fears.

The hypothesised role of a reduced sensitivity to evidence diversity in the maintenance of OCD symptoms could be examined in future experimental work. For example, consider an exposure-based paradigm that aims to first establish and then extinguish fear of contamination via accidental contact with 'dirty' substances (e.g., mud, dirt, stains, blood). Self-report and physiological measures could be collected to track fear reduction as extinction proceeds via the presentation of diverse or non-diverse sets of safety information. It would also be expected that after exposure to a variety of 'safe' episodes, individuals who are less sensitive to evidence diversity may not generalise this experience to other cases of perceived contamination.

Distorted Perception of Similarity

Generally, the degree to which fears are generalised is dependent on the degree to which the conditioned stimulus is perceived as similar to the new stimulus (Boyle, Roche, Dymond, & Hermans, in press), with greatest generalisation occurring for the stimuli most similar to the one that has been conditioned to trigger a fear response (Dunsmoor & Murphy, 2014). In animal conditioning research, the degree of relatedness between stimuli is often manipulated by physical degree: for example, gradients of colours (Dunsmoor et al., 2012) or frequency of tones (Dunsmoor & Murphy, 2014).

Fear generalisation in humans is more complex, due to a unique ability to impose interpretations on the perceptual and contextual details of the learned event (Dunsmoor & Murphy, 2015), and to be influenced by beliefs about relationships between stimuli (Ahmed & Lovibond, 2015, in press). Indeed, recent research has demonstrated that inductive sampling heuristics can influence how one generalises fears, with learnt fear more readily generalised to typical category members relative to atypical members (Dunsmoor & Murphy, 2014). Nevertheless, recent research has reaffirmed that the degree to which stimuli are perceptually or conceptually similar facilitates human fear generalisation (e.g., Bennett et al., 2015; Dunsmoor et al., 2012; Dunsmoor & Murphy, 2014).

In Experiment 5, individuals high in OCD symptomatology appeared to differ from those low in OCD symptomatology in their perception of premise similarity, more often selecting the previously nominated ‘diverse’ premises as being more similar to one another than the ‘non-diverse’ premises. This was true of both OCD-neutral and OCD-relevant items.

Differences in the way that individuals high in OCD symptoms perceive similarity relative to those low in symptoms may influence the way (and even extent to which) these individuals generalise fears. For example, if one perceives similarities between the handle on the door of a public toilet and the handle of a car door (e.g., focusing on colour and size, whereas others might perceive these handles to be more dissimilar than similar as they focus on shape and location), contamination fears may be more easily transferred from toilet door handles to car door handles. Higher perceived similarity between otherwise dissimilar items increases the likelihood of a broad range of stimuli becoming emotionally significant.

Note, however, that this result was not replicated in a clinical sample, and thus caution must be exercised when interpreting these results. Apart from the obvious issue of reduced statistical power in the clinical patient experiment, it is not immediately clear what might explain the failure to replicate this difference. Given the crucial role that perceived similarity plays in clinical conditions (e.g., fear generalisation), it will be important for future work to re-examine perceptions of similarity in a larger sample.

Intact Sensitivity to Sample Size

Both individuals high in OCD symptoms and OCD patients were no poorer at using sample size information to make their inductive judgments than those low in symptoms and non-clinical controls, respectively. This indicates that individuals high in OCD symptomatology understand that an increasing number of positive examples should increase argument strength. Further, this implies that the well-established OCD patient hesitancy and delayed decision-making on other tasks of reasoning (e.g., Fear & Healy, 1997; Foa et al., 2003; Milner et al., 1971) are likely due to patients' higher thresholds for certainty, rather than their misunderstanding of the contributions of positive evidence. This is supported by the notion that, although individuals high in OCD symptomatology demonstrated intact sensitivity to sample size in Experiment 1, they were less confident in their argument evaluations compared to those low in symptoms.

Implications for Cognitive Therapy Approaches to Obsessive-Compulsive Disorder

Collectively, the results presented in this thesis imply that individuals high in OCD symptomatology may (at times) have difficulty drawing inferences from given evidence, both when using logical rules and inductive sampling heuristics. In addition to potentially being an important source of cognitive distortion, this difficulty could also be a barrier to cognitive therapy (Foa & Kozak, 1985). Therapies that aim to change beliefs by presenting individuals with relevant evidence (e.g., Beck & Dozois, 2011; Salkovskis, 1985; Salkovskis & Kirk, 1989) should take into account these limitations in reasoning.

Specifically, if reasoning deficits do play a causal role in the maintenance of OCD symptoms, then two important general modifications to the therapeutic recommendations of cognitive therapy models of OCD are suggested. First, some forms of inductive reasoning (e.g., seeing the implications of diverse evidence) in OCD are more in need of remediation than others. Remediation of an understanding of evidence diversity might be done via direct instruction, as training adults to attend to the statistical features of information is effective at increasing the rate at which these features are incorporated into inductive inferences (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Nisbett, Fong, Lehman, & Cheng, 1987; Nisbett, Krantz, Jepson, & Kunda, 1983). Further, children as young as 5 years old have been successfully taught to attend to the diversity of samples when making inductive inferences (Rhodes, Gelman, & Brickman, 2010).

Second, not only should inductive inferences be a target of cognitive therapy, but so too should possible invalid deductive inferences. A reduced ability to distinguish between valid and invalid arguments may maintain obsessions (e.g., patients may have difficulty deciding whether an intrusive thought has any degree of validity), and may further interfere with the derivation of logical conclusions during cognitive therapy. Deductive training may be conducted by simply presenting patients with deductive tasks to complete, or via direct instruction. Simple practice of syllogistic tasks without any feedback on the accuracy of performance can effect an improvement in deductive performance (Johnson-Laird & Steedman, 1978), as can teaching people to distinguish between valid and invalid argument forms (Hatzikiriakou & Metallidou, 2007; Leighton, 2006; I. T. Nelson, Ratliff, Steinhoff, & Mitchell, 2003).

As both inductive and deductive reasoning performance can be improved by direct instruction, this suggests that pre-intervention courses educating OCD patients regarding the implications of diverse evidence and how to distinguish between valid and invalid arguments may facilitate the effectiveness of cognitive therapy. Further, as deficits appear to be global across both OCD-neutral and OCD-relevant content, attempts to retrain inductive and deductive reasoning may be as effective (and less anxiety-provoking for patients) if training begins with OCD-neutral content.

It will also be important to determine whether individuals with poorer deductive and inductive reasoning abilities also experience worse therapeutic outcomes in cognitive therapy. Such a result would strongly imply the need for pre-intervention trials and the integration of psychoeducation regarding reasoning deficits throughout cognitive therapy.

Limitations of the Current Experiments and Directions for Future Research

The Generation of Alternatives and Logical Reasoning

The IBA model of OCD (O'Connor et al., 2005) postulates that part of the difficulty that OCD patients experience when reasoning inductively relates to their tendency to generate distant possibilities on the basis of irrelevant associations. Patients are hypothesised to fixate on these possibilities, rather than dismissing them on the basis of factual or observable evidence. Whilst the current thesis has sought to examine deduction and induction in individuals high in OCD symptomatology, the focus of these examinations has been on well-established deductive and inductive mechanisms. As such, these examinations may be seen as somewhat removed from the original conceptualisation of inductive reasoning being impaired by a process of excessive generalisation of unrelated alternative possibilities.

Future research might seek to examine whether the generation of “distant alternatives” interferes with inductive performance. This might be achieved through the use of an ‘open-ended’ induction task like that developed by Coley and Vasilyeva (2010). In this task, rather than evaluating the inductive plausibility of arguments composed of premises and conclusions supplied by the experimenter, participants are given premise information and are invited to generate their own plausible conclusions (e.g., ‘Plugged in irons have Property X. Plugged in hairdryers also have Property X. What else do you think would have Property X, and why?’). Open-ended induction tasks have high ecological validity, allowing for an examination of how people *construct* inductive inferences, and allowing participants to draw on any knowledge they perceive as relevant.

One prediction that might follow from the assertion that OCD patients generate and are influenced by irrelevant associations when reasoning (O'Connor et al., 2005; Pélissier & O'Connor, 2002) is that if given this open-ended induction task, OCD patients should generate a larger number and range of possible conclusions than low OCD symptom controls. This would represent an important further test of the IBA conceptualisation of inductive performance in OCD.

Clinical Specificity of Reasoning Deficits

The current research program uncovered interesting differences between those high and low in OCD symptomatology in some forms of inductive and deductive reasoning. However, it remains to be established whether these differences reflect a specific impairment associated with OCD rather than some transdiagnostic impairment.

Inferences about this issue based on the current results are limited because of the absence of clinical control groups.

This is particularly important given that the OCI-R was not the only measure associated with reasoning. For example, depression scores on the DASS exhibited a negative correlation with the tendency to give preference to diverse over non-diverse inductive arguments, $r(48) = -.36$, $p = .01$. Thus, recruitment of depressed and anxious control groups is important in future research to determine whether reasoning differences observed between the OCD patients and non-clinical controls are common to all psychological disorders, or are instead specific to the clinical profile of OCD. Learning more about which psychological difficulties (OCD, anxious or depressive symptoms) are more predictive of reasoning impairments will better inform future possible training approaches.

Alternative Theoretical Frameworks

In recent years, there has been a growing interest in Bayesian models of induction (e.g., Kemp & Tenenbaum, 2009; Oaksford & Hahn, 2007). These models of induction might be applied to examinations of differences in inductive reasoning between clinical and non-clinical groups of participants, and to examinations of individual differences in induction more generally. For example, a Bayesian approach to induction would allow for detailed analyses of whether inductive differences between individuals high and low in OCD symptoms reflect: a) differences in prior beliefs about the risks presented in OCD-relevant scenarios (e.g., contamination, safety), and/or b) differences in the updating of beliefs in the light of evidence accumulation. A Bayesian analysis would be particularly useful in tasks that allow for beliefs to be updated on a trial by trial basis (e.g., the balls and urns task, see Fear & Healy, 1997).

Concluding Statements

In contrast to the “spared deduction, impaired induction account of OCD”, the results of the current thesis have demonstrated that the deductive reasoning abilities of those high in OCD symptoms can be impaired on class-inclusion validity tasks. Those high in OCD symptoms were well able to employ background causal knowledge and amounts of positive evidence when making inductive judgments, but were less likely than controls to use evidence diversity as a basis for property generalisation. Individuals high in OCD symptoms also seemed to perceive similarity between premises somewhat differently to individuals low in symptoms.

These results suggest that the pattern of “spared” and “impaired” reasoning in OCD on inductive and deductive tasks is considerably more complex than has been suggested by models such as the Inference-Based Approach. Nevertheless, impaired reasoning can be seen as having important implications for the maintenance and treatment of OCD symptoms. Future research is needed to determine whether these reasoning deficits are specific to an OCD population, and whether these deficits play a causal role in aetiology and maintenance of OCD.

References

Aardema, F., O'Connor, K. P., Pélissier, M.-C., & Lavoie, M. E. (2009). The quantification of doubt in obsessive-compulsive disorder. *International Journal of Cognitive Therapy, 2*, 188-205.

Abramowitz, J. S. (2006). Psychometric properties and construct validity of the Obsessive-Compulsive Inventory-Revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders, 20*, 1016-1035.

Abramowitz, J. S., Fabricant, L. E., Taylor, S., Deacon, B. J., McKay, D., & Storch, E. A. (2014). The relevance of analogue studies for understanding obsessions and compulsions. *Clinical Psychology Review, 34*, 206-217.

Abramowitz, J. S., Franklin, M. E., Schwartz, S. A., & Furr, J. M. (2003). Symptom presentation and outcome of cognitive-behavioral therapy for obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology, 71*, 1049-1057.

Abramowitz, J. S., Tolin, D. F., & Diefenbach, G. J. (2005). Measuring change in OCD: Sensitivity of the Obsessive-Compulsive Inventory-Revised. *Journal of Psychopathology and Behavioral Assessment, 27*, 317-324.

Ahmed, O., & Lovibond, P. F. (2015). The impact of previously learned feature-relevance on generalisation of conditioned fear in humans. *Journal of Behavior Therapy and Experimental Psychiatry, 46*, 59-65.

Ahmed, O., & Lovibond, P. F. (in press). The impact of instructions on generalization of conditioned fear in humans. *Behavior Therapy*.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219-235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*, 569-576.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research* (Vol. 2, pp. 231-256): Greenwich, CT: JAI Press.

Antony, M. M., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*, 176-181.

Bannon, S., Gonsalvez, C. J., Croft, R. J., & Boyce, P. M. (2002). Response inhibition deficits in obsessive-compulsive disorder. *Psychiatry Research, 110*, 165-174.

Beck, A. T., & Dozois, D. J. A. (2011). Cognitive therapy: Current status and future directions. *Annual Review of Medicine, 62*, 397-409.

Beck, A. T., & Steer, R. A. (1987). Beck Depression Inventory. San Antonio, TX: The Psychological Corporation.

Bédard, M.-J., Joyal, C., Godbout, L., & Chantal, S. (2009). Executive functions and the obsessive-compulsive disorder: On the importance of subclinical symptoms and other concomitant factors. *Archives of Clinical Neuropsychology, 24,* 585-598.

Beggs, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology, 81*, 351-354.

Belayachi, S., & Van der Linden, M. (2010). Feeling of doing in obsessive-compulsive checking. *Consciousness and Cognition, 19*, 534-546.

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*, 25-51.

Bennett, M., Vervoort, E., Boddez, Y., Hermans, D., & Baeyens, F. (2015). Perceptual and conceptual similarities facilitate the generalization of instructed fear. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 149-155.

Berg, E. A. (1948). A simple objective test for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15-22.

Berry, L.-M., & Laskey, B. (2012). A review of obsessive intrusive thoughts in the general population. *Journal of Obsessive-Compulsive and Related Disorders, 1*, 125-132.

Boone, K. B., Ananth, J., Philpott, L., Kaur, A., & Djenderedjian, A. (1991). Neuropsychological characteristics of nondepressed adults with obsessive-compulsive disorder. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 4*, 96-109.

Boschen, M. J., & Vuksanovic, D. (2007). Deteriorating memory confidence, responsibility perceptions and repeated checking: Comparisons in OCD and control samples. *Behaviour Research and Therapy, 45*, 2098-2109.

Boyle, S., Roche, B., Dymond, S., & Hermans, D. (in press). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion*, 1-13.

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182-203.

Braine, M. D. S., & O'Brien, D. P. (1998). The theory of mental-propositional logic: Description and illustration. In M. D. S. Braine (Ed.), *Mental logic* (pp. 79-90). Mahwah, NJ: Lawrence Erlbaum Associates.

Bright, A., & Feeney, A. (2014). The engine of thought is a hybrid: Roles of associative and structured knowledge in reasoning. *Journal of Experimental Psychology: General, 143*, 2082-2102.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Brown, T. A., & Barlow, D. H. (2014). *Anxiety and related disorders interview schedule for DSM-5 (ADIS-5): Client interview schedule*. New York, NY: Oxford University Press.

Bucci, P., Galderisi, S., Catapano, F., Di Benedetto, R., Piegari, G., Mucci, A., & Maj, M. (2007). Neurocognitive indicies of executive hypercontrol in obsessive-compulsive disorder. *Acta Psychiatrica Scandinavica, 115*, 380-387.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5.

Burns, G. L., Formea, G. M., Keortge, S., & Sternberger, L. G. (1995). The utilization of nonpatient samples in the study of obsessive compulsive disorder. *Behaviour Research and Therapy, 33*, 133-144.

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders, 21*, 105-117.

Carr, A. T. (1974). Compulsive neurosis: A review of the literature. *Psychological Bulletin, 81*, 311-318.

Chamberlain, S. R., Blackwell, A. D., Fineberg, N. A., Robbins, T. W., & Sahakian, B. J. (2005). The neuropsychology of obsessive compulsive disorder: The importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neuroscience and Biobehavioral Reviews, 29*, 399-419.

Choi, I., Nisbett, R. E., & Smith, E. E. (1997). Culture, category salience, and inductive reasoning. *Cognition, 65*, 15-32.

Clark, D. A. (2005). Lumping versus splitting: A commentary on subtyping in OCD. *Behavior Therapy, 36*, 401-404.

Clark, D. M. (1986). A cognitive approach to panic. *Behaviour Research and Therapy, 24*, 461-470.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope & F. R. Schneier (Eds.), *Social phobia* (pp. 69-93). New York, NY: Guilford.

Coles, M. E., Radomsky, A. S., & Horng, B. (2006). Exploring the boundaries of memory distrust from repeated checking: Increasing external validity and examining thresholds. *Behaviour Research and Therapy, 44*, 995-1006.

Coley, J. D., & Vasilyeva, N. Y. (2010). Generating inductive inferences: Premise relations and property effects. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 184-226). San Diego, CA: Elsevier Academic Press.

Constans, J. I., Foa, E. B., Franklin, M. E., & Mathews, A. (1995). Memory for actual and imagined events in OC checkers. *Behaviour Research and Therapy, 33*, 665-671.

Cooksey, R. W., & Freebody, P. (1987). Aspects of a computer-managed test of children's reading vocabulary: Reliability, validity and characterization of knowledge. *Reading Psychology, 8*, 103-118.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLos ONE, 8*, e57410.

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 382-392.

de Jong, P. J., Weertman, A., Horselenberg, R., & van den Hout, M. A. (1997). Deductive reasoning and pathological anxiety: Evidence for a relatively strong "belief bias" in phobic subjects. *Cognitive Therapy and Research, 21*, 647-662.

Deckersbach, T., Otto, M. W., Savage, C. R., Baer, L., & Jenike, M. A. (2000). The relationship between semantic organization and memory in obsessive-compulsive disorder. *Psychotherapy and Psychosomatics*, *69*, 101-107.

Dickstein, L. S. (1975). Effects of instructions and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology, 77*, 514-516.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory & Cognition, 6*, 76-83.

Dickstein, L. S. (1981). Conversion and possibility in syllogistic reasoning. *Bulletin of the Psychonomic Society, 18*, 229-232.

Dube, C. (2009). *Dual process theory and syllogistic reasoning: A signal detection analysis* (Doctoral dissertation, University of Massachusetts). Retrieved from http://scholarworks.umass.edu/

Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review, 117*, 831-863.

Dugas, M. J., Freeston, M. H., & Ladouceur, R. (1997). Intolerance of uncertainty and problem orientation in worry. *Cognitive Therapy and Research, 21*, 593-606.

Dugas, M. J., Heydayati, M., Karavisdas, A., Buhr, K., Francis, K., & Phillips, N. A. (2005). Intolerance of uncertainty and information processing: Evidence of biased recall and interpretations. *Cognitive Therapy and Research, 29,* 57-70.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*, 91-101.

Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology, 89*, 300-305.

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25*, 1816-1821.

Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: How humans generalize fear. *Trends in Cognitive Sciences, 19*, 73-77.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 37,* 809-829.

Evans, J. S. B. T. (1982). *The psychology of deductive reasoning*. London, UK: Routledge & Kegan Paul.

Evans, J. S. B. T. (1989). *Biases in human reasoning: Causes and consequences*. London: Erlbaum.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255-278.

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*, 295-306.

Evans, J. S. B. T., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54A*, 935-958.

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition, 35,* 1772-1784.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Evans, J. S. B. T., & Over, D. E. (2013). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking & Reasoning, 19*, 267-283.

Evans, J. S. B. T., Over, D. E., & Manktelow, K. I. (1993). Reasoning, decision making and rationality. *Cognition, 49*, 165-187.

Falmagne, R. J., & Gonsalves, J. (1995). Deductive inference. *Annual Review of Psychology, 46*, 525-559.

Farrell, L., Barrett, P., & Piacentini, J. (2006). Obsessive-compulsive disorder across the developmental trajectory: Clinical correlates in children, adolescents and adults. *Behaviour Change, 23*, 103-120.

Fear, C. F., & Healy, D. (1997). Probabilistic reasoning in obsessive-compulsive and delusional disorders. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 27*, 199-208.

Feeney, A. (2007). How many processes underlie category-based induction? Effects of conclusion specificity and cognitive ability. *Memory & Cognition, 35*, 1830-1839.

Feeney, A., & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning, 17*, 156-181.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment, 14*, 485-495.

Foa, E. B., Illai, D., McCarthy, P. R., Shoyer, B., & Murdock, T. (1993). Information processing in obsessive-compulsive disorder. *Cognitive Therapy and Research, 17*, 173-189.

Foa, E. B., & Kozak, M. J. (1985). Treatment of anxiety disorders: Implications for psychopathology. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 421-452). Hillsdale, NJ: Erlbaum.

Foa, E. B., & Kozak, M. J. (1995). DSM-IV field trial: Obsessive-compulsive disorder. *American Journal of Psychiatry, 152*, 90-96.

Foa, E. B., Mathews, A., Abramowitz, J. S., Amir, N., Przeworski, A., Riggs, D. S., . . . Alley, A. (2003). Do patients with obsessive-compulsive disorder have deficits in decision-making? *Cognitive Therapy and Research, 27*, 431-445.

Foa, E. B., & McNally, R. J. (1986). Sensitivity to feared stimuli in obsessive-compulsives: A dichotic listening analysis. *Cognitive Therapy and Research, 10*, 477-485.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253-292.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120*, 34-45.

Geller, D. A., Biederman, J., Faraone, S., Agranat, A., Cradock, K., Hagermoser, L., . . . Coffey, B. (2001). Developmental aspects of obsessive compulsive disorder: Findings in children, adolescents, and adults. *The Journal of Nervous and Mental Disease, 189*, 471-477.

Gibbs, N. A. (1996). Nonclinical populations in research on obsessive-compulsive disorder: A critical review. *Clinical Psychology Review, 16*, 729-773.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason: A localization study of deductive and inductive reasoning using PET (O15) blood flow technique. *NeuroReport, 8*, 1305-1310.

Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 13-36). New York, NY: Cambridge University Press.

Gönner, S., Leonhart, R., & Ecker, W. (2008). The Obsessive-Compulsive Inventory-Revised (OCI-R): Validation of the German Version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders, 22*, 734-749.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*, 213-224.

Grant, D. A., & Berg, E. A. (1948). 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.

Greene, T. R. (1989). Children's understanding of class-inclusion hierarchies: The relationship between external representation and task performance. *Journal of Experimental Child Psychology, 48*, 62-89.

Guidano, V. F., & Liotti, G. (1983). *Cognitive processes and emotional disorders*. New York, NY: Guilford Press.

Guntheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64*, 159-174.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition, 33*, 289-302.

Hájek, A., & Hall, N. (2002). Induction and probability. In P. Machamer & M. Silberstein (Eds.), *The Blackwell guide to the philosphy of science* (pp. 149-172). Malden, MA: Blackwell Publishers.

Halford, G. S., Andrews, G., & Jensen, I. (2002). Integration of category induction and hierarchical classification: One paradigm at two levels of complexity. *Journal of Cognition and Development, 3*, 143-177.

Harman, G. (1999). *Reasoning, meaning, and mind*. Oxford, UK: Oxford University Press.

Hatzikiriakou, K., & Metallidou, P. (2007). Teaching deductive reasoning to pre-service teachers: Promises and constraints. *International Journal of Science and Mathematics Education, 7*, 81-101.

Hayes, B. K. (2007). The development of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches*. New York, NY: Cambridge University Press.

Hayes, B. K., Goodhew, A., Heit, E., & Gillan, J. (2003). The role of diverse instruction in conceptual change. *Journal of Experimental Child Psychology, 86*, 253-276.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Cognitive Science, 1*, 278-292.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248-274). Oxford, UK: Oxford University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*, 569-592.

Heit, E. (2007). What is induction and why study it? In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 1-24). Cambridge, UK: Cambridge University Press.

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43*, 243-273.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In W. Ahn, B. C. Goldstone, A. B. Love & P. Wolff (Eds.), *Categorization inside and outside the lab: Festscrift in honor of Douglas L. Medin* (pp. 87-100). Washington, DC: American Psychological Association.

Heit, E., & Rotello, C. M. (2005). Are there two kinds of reasoning? In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual meeting of the Cognitive Science Society* (pp. 923-928). Mahwah, NJ: Erlbaum.

Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 805-812.

Heit, E., & Rotello, C. M. (2012). The pervasive effects of argument length on inductive reasoning. *Thinking & Reasoning, 18*, 244-277.

Heit, E., Rotello, C. M., & Hayes, B. K. (2012). Relations between memory and reasoning. *Psychology of Learning and Motivation - Advances in Research and Theory, 57*, 57-101.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 411-422.

Hermans, D., Martens, K., De Cort, K., Pieters, G., & Eelen, P. (2003). Reality monitoring and metacognitive beliefs related to cognitive confidence in obsessive-compulsive disorder. *Behaviour Research and Therapy, 41*, 383-401.

Holaway, R. M., Heimberg, R. G., & Coles, M. E. (2006). A comparison of intolerance of uncertainty in analogue obsessive-compulsive disorder and generalized anxiety disorder. *Anxiety Disorders, 20*, 158-174.

Holyoak, K. J., & Morrison, R. G. (2005). Thinking and reasoning: A reader's guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. New York, NY: Cambridge University Press.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14,* 399-425.

Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd ed.). New York, NY: McGraw-Hill.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50,* 189-209.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50,* 109-135.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62-88.

Johnson-Laird, P. N., Mancini, F., & Gangemi, A. (2006). A hyper-emotion theory of psychological illnesses. *Psychological Review, 113*, 822-841.

Johnson-Laird, P. N., & Shafir, E. (1993). The interaction between reasoning and decision making: An introduction. *Cognition, 49,* 1-9.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology, 10*, 64-99.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman

(Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York, NY: Cambridge University Press.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn & Bacon.

Kaufman, J. C., & Kaufman, A. S. (2001). Time for the changing of the guard: A farewell to short forms of intelligence tests. *Journal of Psychoeducational Assessment, 19,* 245-267.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57,* 227-254.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116,* 20-58.

Kemp, R., Chua, S., McKenna, P., & David, A. (1997). Reasoning and delusions. *British Journal of Psychiatry, 170,* 398-405.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 593-602.

Khawaja, N. G., & Yu, L. (2010). A comparison of the 27-item and 12-item Intolerance of Uncertainty Scales. *Clinical Psychologist, 14*, 97-106.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition, 31*, 155-165.

Kincannon, A., & Spellman, B. A. (2003). The use of category and similarity information in limiting hypotheses. *Memory & Cognition, 31*, 114-132.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852-884.

Koran, L. M., Thienemann, M. L., & Davenport, R. (1996). Quality of life for patients with obsessive-compulsive disorder. *American Journal of Psychiatry, 153,* 783-788.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797-2822.

Kozak, M. J., Foa, E. B., & McCarthy, P. R. (1987). Obsessive-compulsive disorder. In C. Last & M. Hersen (Eds.), *Handbook of anxiety disorders* (Vols. 87-108). New York, NY: Pergamon Press.

Ladouceur, R., Talbot, F., & Dugas, M. J. (1997). Behavioral expressions of intolerance of uncertainty in worry. *Behavior Modification, 21*, 355-371.

Lassiter, D., & Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition, 136*, 123-134.

Lavy, E., van Oppen, P., & van den Hout, M. A. (1994). Selective processing of emotional information in obsessive compulsive disorder. *Behaviour Research and Therapy, 32*, 243-246.

Lawrence, N. S., Wooderson, S., Mataix-Cols, D., David, R., Speckens, A., & Phillips, M. L. (2006). Decision making and set shifting impairments are associated with distinct symptom dimensions in obsessive-compulsive disorder. *Neuropsychology, 20*, 409-419.

Leighton, J. (2006). Teaching and assessing deductive reasoning skills. *The Journal of Experimental Education, 74*, 109-136.

Leon, A. C., Portera, L., & Weissman, M. M. (1995). The social costs of anxiety disorders. *British Journal of Psychiatry, 27*, 19-22.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. N. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26*, 181-206.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61,* 303-332.

López, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition, 23*, 374-382.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251-291.

López, A., Gelman, S. A., Guntheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development*, *63*, 1070-1090.

Lorian, C. N., & Grisham, J. R. (2011). Clinical implications of risk aversion: An online study of risk-avoidance and treatment utilization in pathological anxiety. *Journal of Anxiety Disorders, 25*, 840-848.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney: Psychology Foundation.

Lucey, J. V., Burness, C. E., Costa, D. C., Gacinovic, S., Pilowsky, L. S., Ell, P. J., . . . Kerwsin, R. W. (1997). Wisconsin Card Sorting Task (WCST) errors and cerebral blood flow in obsessive-compulsive disorder (OCD). *British Journal of Medical Psychology, 70,* 403-411.

MacDonald, P. A., Antony, M. M., MacLeod, C. M., & Richter, M. M. (1997). Memory and confidence in memory judgments among individuals with obsessive compulsive disorder and non-clinical controls. *Behaviour Research and Therapy*, *35*, 497-505.

Mataix-Cols, D., Rosario-Campos, M. C., & Leckman, J. F. (2005). A multidimensional model of obsessive-compulsive disorder. *American Journal of Psychiatry, 162,* 228-238.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. *Cognition, 59*, 199-217.

McEvoy, P. M., Grove, R., & Slade, T. (2011). Epidemiology of anxiety disorders in the Australian general population: Findings of the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Psychiatry, 45*, 957-967.

McGuire, L., Junginger, J., Adams, S. G. Jr., Burright, R., & Donovick, P. (2001). Delusions and delusional reasoning. *Journal of Abnormal Psychology, 110*, 259-266.

McNally, R. J. (1999). Information-processing abnormalities in obsessive-compulsive disorder. In W. Goodman, M. Rudorfer & J. Maser (Eds.), *Obsessive-compulsive disorder: Contemporary issues in treatment*. Mahway, NJ: Lawrence Erlbaum.

McNally, R. J., Amir, N., Louro, C. E., Lukach, B. M., Riemann, B. C., & Calamari, J. E. (1994). Cognitive processing of idiographic emotional information in panic disorder. *Behaviour Research and Therapy, 32*, 119-122.

McNally, R. J., & Kohlbeck, P. A. (1993). Reality monitoring in obsessive-compulsive disorder. *Behaviour Research and Therapy, 31*, 249-253.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517-532.

Milner, A. D., Beech, H. R., & Walker, V. J. (1971). Decision processes and obsessional behaviour. *British Journal of Social and Clinical Psychology, 10*, 88-89.

Moritz, S., Kloss, M., & Jelinek, L. (2010). Negative priming (cognitive inhibition) in obsessive-compulsive disorder (OCD). *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 1-5.

Muller, J., & Roberts, J. E. (2005). Memory and attention in obsessive-compulsive disorder: A review. *Anxiety Disorders, 19*, 1-28.

Myers, S. G., & Wells, A. (2013). An experimental manipulation of metacognition: A test of the metacognitive model of obsessive-compulsive symptoms. *Behaviour Research and Therapy, 51*, 177-184.

Nakao, T., Nakagawa, A., Nakatani, E., Nabeyama, M., Sanematsu, H., Yoshiura, T., ... Kanba, S. (2009). Working memory dysfunction in obsessive-compulsive disorder: A neuropsychological and functional MRI study. *Journal of Psychiatric Research, 43*, 784-791.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353-383.

Nelson, H. E., & Willison, J. R. (1991). *National Adult Reading Test* (2nd ed.). Windsor, UK: NFER-NELSON Publishing Company.

Nelson, I. T., Ratliff, R. L., Steinhoff, G., & Mitchell, G. J. (2003). Teaching logic to auditing students: Can training in logic reduce audit judgment errors? *Journal of Accounting Education, 21*, 215-237.

Newstead, S. E., & Evans, J. S. B. T. (1993). Mental models as an explanation of belief bias effects in syllogistic reasoning. *Cognition, 46*, 93-97.

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. (1992). The source of belief bias in syllogistic reasoning. *Cognition, 45*, 257-284.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625-631.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday reasoning. *Psychological Review, 90*, 339-363.

O'Connor, K. P., Aardema, F., & Pélissier, M.-C. (2005). *Beyond reasonable doubt: Reasoning processes in obsessive-compulsive disorder and related disorders*. New York, NY: John Wiley & Sons Ltd.

O'Connor, K. P., & Robillard, S. (1995). Inference processes in obsessive-compulsive disorder: Some clinical observations. *Behaviour Research and Therapy, 33*, 887-896.

O'Connor, K. P., & Robillard, S. (1999). A cognitive approach to the treatment of primary inferences in obsessive-compulsive disorder. *Journal of Cognitive Psychotherapy, 13*, 359-375.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608-631.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*, 349-357.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Oaksford, M., & Hahn, U. (2007). Induction, deduction and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental and computational approaches* (pp. 269-301). Cambridge, UK: Cambridge University Press.

Obsessive Compulsive Cognitions Working Group. (1997). Cognitive assessment of obsessive-compulsive disorder. *Behaviour Research and Therapy, 37*, 667-681.

Okasha, A., Rafaat, M., Mahallawy, N., El Nahas, G., Seif El Dawla, A., Sayed, M., & El Kholi, S. (2000). Cognitive dysfunction in obsessive-compulsive disorder. *Acta Psychiatrica Scandinavica, 101*, 281-285.

Ólafsson, R. P., Snorrason, Í., Bjarnason, R. K., Emmelkamp, P. M. G., Ólason, D. ó., & Kristjánsson, Á. (2014). Replacing intrusive thoughts: Investigating thought control in relation to OCD symptoms. *Journal of Behavior Therapy and Experimental Psychiatry, 45*, 506-515.

Olley, A., Malhi, G., & Sachdev, P. (2007). Memory and executive functioning in obsessive-compulsive disorder: A selective review. *Journal of Affective Disorders, 104*, 15-23.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*, 237-241.

Osherson, D. N., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive verses probabilistic reasoning. *Neuropsychologica, 36*, 369-376.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185-200.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184-188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Pearsons, J. B., & Foa, E. B. (1984). Processing of fearful and neutral information by obsessive-compulsives. *Behaviour Research and Therapy, 22*, 259-265.

Peer, E., Vosgerau, J., & Acquisiti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*, 1023-1031.

Peirce, C. S. (1901). On the logical of drawing history from ancient documents especially from testimonies (logic of history). *Collected Papers v. 7*, paragraph 219.

Pélissier, M.-C., & O'Connor, K. P. (2002). Deductive and inductive reasoning in obsessive-compulsive disorder. *British Journal of Clinical Psychology, 41*, 15-27.

Pélissier, M.-C., O'Connor, K. P., & Dupuis, G. (2009). When doubting begins: Exploring inductive reasoning in obsessive-compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 39-49.

Penadés, R., Catalán, R., Rubia, K., Andrés, S., Salamero, M., & Gastó, C. (2007). Impaired response inhibition in obsessive compulsive disorder. *European Psychiatry, 22,* 404-410.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102*, 533-566.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 811-828.

Purdon, C., & Clark, D. A. (1993). Obsessive intrusive thoughts in non-clinical subjects. Part I. Content and relation with depressive, anxious, and obsessional symptoms. *Behaviour Research and Therapy, 31*, 713-720.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 53A*, 1202-1223.

Rachman, S. J. (1997). A cognitive theory of obsessions. *Behaviour Research and Therapy, 35*, 793-802.

Rachman, S. J. (1998). A cognitive theory of obsessions: Elaborations. *Behaviour Research and Therapy, 36*, 385-401.

Rachman, S. J., & de Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy, 16*, 233-248.

Radomsky, A. S., Gilchrist, P. T., & Dassault, D. (2006). Repeated checking does cause memory distrust. *Behaviour Research and Therapy, 44*, 305-316.

Rankins, D., Bradshaw, J. L., & Georgiou-Karistianis, N. (2005). Local-global processing in obsessive-compulsive disorder and comorbid Tourette's syndrome. *Brain and Cognition, 59*, 43-51.

Rasmussen, S. A., & Eisen, J. L. (1992). The epidemiology and differential diagnosis of obsessive compulsive disorder. *Journal of Clinical Psychiatry, 53*, 4-10.

Rasmussen, S. A., & Tsuang, M. T. (1986). DSM-III obsessive compulsive disorder: Clinical characteristics and family history. *The American Journal of Psychiatry, 143*, 317-322.

Rassin, E., Cougle, J. R., & Muris, P. (2007). Content difference between normal and abnormal obsessions. *Behaviour Research and Therapy, 45,* 2800-2803.

Rassin, E., & Muris, P. (2007). Abnormal and normal obsessions: A reconsideration. *Behaviour Research and Therapy, 45,* 1065-1070.

Reed, G. F. (1997). Obsessional cognition: Performance on two numerical tasks. *British Journal of Psychiatry, 130*, 184-185.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition, 34,* 3-16.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science, 33,* 301-343.

Rehder, B., & Burnett, R. C. (2005). Feature interference and the causal structure of categories. *Cognitive Psychology, 50*, 364-374.

Revlin, R., Leirer, V., Yopp, H., & Yopp, R. (1980). The belief bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition, 8,* 584-592.

Rhodes, M., Gelman, S. A., & Brickman, D. (2008). Developmental changes in the consideration of sample diversity in inductive reasoning. *Journal of Cognition and Development*, *9*, 112-143.

Rhodes, M., Gelman, S. A., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science, 13,* 421-429.

Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12,* 129-134.

Roh, K. S., Shin, M. S., Kim, M.-S., Lee, K. J., & Kwon, J. S. (2005). Persistent cognitive dysfunction in patients with obsessive-compulsive disorder: A naturalistic study. *Psychiatry and Clinical Neurosciences, 59*, 539-545.

Rotello, C. M., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1317-1330.

Rotge, J.-Y., Langbour, N., Dilharreguy, B., Bordessoules, M., Guehl, D., Bioulac, B., . . . Burbaud, P. (2013). Contextual and behavioral influences on uncertainty in obsessive-compulsive disorder. *Cortex, 23*, 1-10.

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry, 15*, 53-63.

Salkovskis, P. M. (1985). Obsessional-compulsive problems: A cognitive-behavioural analysis. *Behaviour Research and Therapy, 23*, 571-583.

Salkovskis, P. M. (1989). Cognitive-behavioural factors and the persistence of intrusive thoughts in obsessional problems. *Behaviour Research and Therapy, 27*, 677-682.

Salkovskis, P. M. (1999). Understanding and treating obsessive–compulsive disorder. *Behaviour Research and Therapy, 37*, S29-S52.

Salkovskis, P. M., & Harrison, J. (1984). Abnormal and normal obsessions: A replication. *Behaviour Research and Therapy, 22*, 549-552.

Salkovskis, P. M., & Kirk, J. (1989). Obsessional disorders. In K. Hawton, P. M. Salkovskis, J. Kirk & D. M. Clark (Eds.), *Cognitive-behavior therapy for psychiatric problems: A practical guide* (pp. 129-168). New York, NY: Oxford University Press.

Sanz, M., Molina, V., Calcedo, A., Martin-Loeches, M., & Rubia, F. J. (2001). The Wisconsin Card Sorting Test and the assessment of frontal function in obsessive-compulsive patients: An event-related potential study. *Cognitive Neuropsychiatry, 6*, 109-129.

Shafran, R., & Rachman, S. J. (2004). Thought-action fusion: A review. *Journal of Behavior Therapy and Experimental Psychiatry, 35*, 87-107.

Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 641-649.

Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review, 14*, 890-894.

Shafto, P., Kemp, C., Bonaqitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition, 109,* 175-192.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1*, 213-220.

Shipley, E. F., & Shepperson, B. (2006). Test sample selection by preschool children: Honoring diversity. *Memory & Cognition, 34*, 1444-1451.

Simpson, J., Cove, J., Fineberg, N., Msetfi, R. M., & Ball, L. J. (2007). Reasoning in people with obsessive-compulsive disorder. *British Journal of Clinical Psychology, 46*, 397-411.

Singmann, H., & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning, 17*, 247-281.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231-280.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7,* 246-251.

Smeets, G., & De Jong, P. J. (2005). Belief bias and symptoms of psychopathology in a non-clinical sample. *Cognitive Therapy and Research, 29*, 377-386.

Soref, A., Dar, R., Argov, G., & Meiran, N. (2008). Obsessive-compulsive tendencies are associated with a focused information processing strategy. *Behaviour Research and Therapy, 46*, 1295-1299.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

Stanovich, K. E., & West, R. F. (1998a). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-188.

Stanovich, K. E., & West, R. F. (1998b). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 26*, 161-179.

Stern, E. R., Welsh, R. C., Gozalez, R., Fitzgerald, K. D., Abelson, J. L., & Taylor, S. F. (2013). Subjective uncertainty and limic hyperactivation in obsessive-compulsive disorder. *Human Brain Mapping, 34*, 1956-1370.

Stevens, J. (1996). Applied multivariate statistics for the social sciences. Mahwah, NJ: Erlbaum.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology, 48A*, 613-643.

Storch, E. A., Merlo, L. J., Lehmkuhl, H., Geffken, G. R., Jacob, M., Ricketts, E., . . . Goodman, W. K. (2008). Cognitive-behavioral therapy for obsessive-compulsive disorder: A non-randomized comparison of intensive and weekly approaches. *Journal of Anxiety Disorders, 22*, 1146-1158.

Tata, P. R., Leibowitz, J. A., Pmnty, M. J., Cameron, M., & Pickering, A. D. (1996). Attentional bias in obsessional compulsive disorder. *Behaviour Research and Therapy, 37*, 161-166.

Taylor, S. (1998). Assessment of obsessive-compulsive disorder. In R. P. Swinson, M. M. Antony, S. Rachman & M. A. Richter (Eds.), *Obsessive-compulsive disorder: Theory, research, and treatment.* New York, NY: Guilford Press.

Taylor, S. (2002). Cognition in obsessive compulsive disorder: An overview. In R. O. Frost & G. Steketee (Eds.), *Cognitive approaches to obsessions and compulsions: Theory, assessment and treatment*. Oxford, UK: Elsevier.

Thompson, V. A. (1996). Reasoning from false premises: The role of soundness in making logical deductions. *Canadian Journal of Experimental Psychology, 50*, 315-319.

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., Amir, N., Street, G. P., & Foa, E. B. (2001). Memory and memory confidence in obsessive-compulsive disorder. *Behaviour Research and Therapy, 39*, 913-927.

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., & Foa, E. B. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *Anxiety Disorders, 17*, 233-242.

van den Hout, M. A., & Kindt, M. (2003a). Phenomenological validity of an OCD-memory model and the remember/know distinction. *Behaviour Research and Therapy, 41*, 369-378.

van den Hout, M. A., & Kindt, M. (2003b). Repeated checking causes memory distrust. *Behaviour Research and Therapy, 46*, 98-113.

Veale, D. M., Sahakian, B. J., Owen, A. M., & Marks, I. M. (1996). Specific cognitive deficits in tests sensitive to frontal lobe dysfunction in obsessive-compulsive disorder. *Psychological Medicine, 26*, 1261-1269.

Volans, P. J. (1976). Styles of decision-making and probability appraisal in selected obsessional and phobic patients. *British Journal of Social & Clinical Psychology, 15*, 305-317.

Vroling, M. S., & de Jong, P. J. (2009). Deductive reasoning and social anxiety: Evidence for a fear-confirming belief bias. *Cognitive Therapy and Research, 33*, 633-644.

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

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135-151). Harmondsworth, UK: Penguin.

Wechsler, D. (1981). *WAIS-R manual*. New York, NY: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Pearson.

Wechsler, D. (2008). *Wechsler adult intelligence scale - Fourth Edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.

Winer, G. A. (1980). Class-inclusion reasoning in children: A review of the empirical literature. *Child Development, 51*, 309-328.

Yovel, I., Revelle, W., & Mineka, S. (2005). Who sees trees before forest? The obsessive-compulsive style of visual attention. *Psychological Science, 16*, 123-129.

Zielinski, T. A., Goodwin, G. P., & Halford, G. S. (2010). Complexity of categorical syllogisms: An integration of two metrics. *European Journal of Cognitive Psychology, 22*, 391-421.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection methods, and some correlates of word and nonword recognition. *Intelligence, 1*, 5-31.

APPENDICES

Appendix A

Chapter 2: Sensitivity to Premise Monotonicity and Argument Validity in Obsessive-Compulsive Disorder

Experiment 1: Item Set

Table A 1.1

Experiment 1 Premise Monotonicity Induction-Deduction Task Item Set

Valid						Invalid					
Premise 1	Premise 2	Premise 3	Premise 4	Premise 5	Conclusion	Premise 1	Premise 2	Premise 3	Premise 4	Premise 5	Conclusion
mammals	-	-	-	-	goats	deer	-	-	-	-	mammals
mammals	-	-	-	-	zebras	lions	-	-	-	-	foxes
mammals	-	-	-	-	monkeys	dogs	-	-	-	-	horses
fruits	-	-	-	-	oranges	cherries	-	-	-	-	fruits
fruits	-	-	-	-	bananas	plums	-	-	-	-	peaches
fruits	-	-	-	-	apricots	strawberries	-	-	-	-	grapes
flowers	-	-	-	-	daisies	daffodils	-	-	-	-	flowers
flowers	-	-	-	-	frangipanis	tulips	-	-	-	-	lavenders
alcohol	-	-	-	-	whiskey	tequila	-	-	-	-	alcohol
alcohol	-	-	-	-	vodka	cognac	-	-	-	-	champagne
mammals	cats	lions	-	-	sheep	tigers	pandas	cows	-	-	mammals
mammals	pigs	zebras	-	-	rats	mice	gorillas	pandas	-	-	rabbits
mammals	zebras	wolves	-	-	pandas	goats	bears	rats	-	-	tigers
fruits	mandarins	mangoes	-	-	blueberries	raspberries	bananas	mandarins	-	-	fruits
fruits	pears	apricots	-	-	strawberries	apples	mangoes	rockmelons	-	-	nectarines
fruits	raspberries	pears	-	-	watermelons	blueberries	cherries	limes	-	-	apricots
flowers	sunflowers	violets	-	-	daffodils	daisies	irises	roses	-	-	flowers
flowers	lilies	carnations	-	-	orchids	frangipanis	poppies	dandelions	-	-	irises
alcohol	beer	white wine	-	-	bourbon	whiskey	scotch	tequila	-	-	alcohol
alcohol	gin	sherry	-	-	beer	port	champagne	bourbon	-	-	rum
mammals	cats	horses	goats	pandas	bears	dogs	deer	gorillas	bears	cows	mammals
mammals	rats	tigers	horses	lions	gorillas	cats	bears	foxes	zebras	mice	horses
mammals	rabbits	cats	monkeys	cows	wolves	deer	monkeys	pandas	foxes	lions	rats
fruits	rockmelons	grapes	raspberries	nectarines	bananas	blueberries	apricots	mandarins	nectarines	pears	fruits
fruits	raspberries	bananas	cherries	oranges	plums	raspberries	oranges	limes	mangoes	strawberries	pineapples
fruits	pineapples	blueberries	mandarins	peaches	limes	pears	rockmelon	plums	apples	limes	watermelons
flowers	violets	lilies	irises	lavenders	sunflowers	frangipani	poppies	irises	tulips	orchids	flowers
flowers	daisies	orchids	poppies	lilacs	lilies	carnations	lavenders	violets	daisies	lilacs	dandelions
alcohol	cognac	gin	champagne	scotch	rum	scotch	red wine	sherry	port	beer	gin
alcohol	white wine	whiskey	sherry	bourbon	cognac	port	white wine	tequila	gin	rum	alcohol

Note. All stimuli were preceded by the word "all" and the conclusion "have/has the property". Properties were a random combination of one letter and one number (e.g., Z3).

Experiment 1: Statistical Summaries

Table A 1.2

Experiment 1 Premise Monotonicity Task ANOVA Summary Table for Proportion of Positive Responses (‘Strong’/ ‘Valid’)

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Group		.00	1	.00	1.30	.26	.01
Reasoning Condition		.23	1	.23	2.22	.14	.01
OCD-Group × Reasoning Condition		.00	1	.00	.02	.88	.00
Error		16.49	160	.10			
Within-Subjects Effects							
Validity		104.05	1	104.05	1120.75	<.001**	.88
Validity × OCD-Group		1.63	1	1.63	17.52	<.001**	.10
Validity × Reasoning Condition		.05	1	.05	.59	.45	.00
Validity × OCD-Group × Reasoning Condition		.00	1	.00	.06	.85	.00
Error (Validity)		14.86	160	.09			
Length	Linear	.00	1	.00	.01	.91	.00
	Quadratic	.31	1	.31	22.68	<.001**	.12
Length × OCD-Group	Linear	.00	1	.00	.02	.90	.00
	Quadratic	.01	1	.01	.54	.47	.00
Length × Reasoning Condition	Linear	.11	1	.11	3.40	.07	.02
	Quadratic	.00	1	.00	.18	.67	.00
Length × OCD-Group × Reasoning Condition	Linear	.00	1	.00	.02	.90	.00
	Quadratic	.03	1	.03	1.81	.18	.01
Error (Length)	Linear	5.28	160	.03			
	Quadratic	2.17	160	.01			
Validity × Length	Linear	2.49	1	2.49	109.95	<.001**	.41
	Quadratic	.14	1	.14	13.23	<.001**	.08
Error (Validity × Length)	Linear	3.62	160	.02			
	Quadratic	.174	160	.01			

Note. **p* < .05, ***p* < .01.

Table A 1.3

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Proportion of Positive Responses ('Strong'/ 'Valid') to Valid Arguments

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Group		1.35	1	1.35	17.53	<.001**	.10
Reasoning Condition		.03	1	.03	.39	.53	.00
OCD-Group × Reasoning Condition		.00	1	.00	.001	.98	.00
Error		12.30	160	.08			
Within-Subjects Effects							
Length	Linear	1.21	1	1.21	46.62	<.001**	.23
	Quadratic	.44	1	.44	31.42	<.001**	.16
Length × OCD-Group	Linear	.00	1	.00	.12	.73	.00
	Quadratic	.05	1	.05	3.31	.07	.02
Length × Reasoning Condition	Linear	.00	1	.00	.13	.72	.00
	Quadratic	.00	1	.00	.04	.84	.00
Length × OCD-Group × Reasoning Condition	Linear	.00	1	.00	.003	.95	.00
	Quadratic	.03	1	.03	2.20	.14	.01
Error (Length)	Linear	4.16	160	.03			
	Quadratic	2.23	160	.01			

Note. *p < .05, **p < .01.

Table A 1.4

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Proportion of Positive Responses ('Strong'/ 'Valid') to Invalid Arguments

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Group		.41	1	.41	3.47	.06	.02
Reasoning Condition		.25	1	.25	2.13	.15	.01
OCD-Group × Reasoning Condition		.01	1	.01	.04	.83	.00
Error		12.39	160	.08			
Within-Subjects Effects							
Length	Linear	1.28	1	1.28	43.15	<.001**	.21
	Quadratic	.02	1	.02	1.49	.22	.01
Length × OCD-Group	Linear	.01	1	.01	.27	.60	.00
	Quadratic	.01	1	.01	.83	.36	.01
Length × Reasoning Condition	Linear	.17	1	.17	5.82	.02*	.04
	Quadratic	.01	1	.01	.85	.36	.01
Length × OCD-Group × Reasoning Condition	Linear	.00	1	.00	.09	.76	.00
	Quadratic	.00	1	.00	.21	.65	.00
Error (Length)	Linear	4.74	160	.03			
	Quadratic	1.69	160	.01			

Note. *p < .05, **p < .01.

Table A 1.5

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Confidence Ratings to Valid Arguments

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Group		6.25	1	6.25	7.75	.006**	.05
Reasoning Condition		.27	1	.27	.34	.56	.00
OCD-Group × Reasoning Condition		.15	1	.15	.18	.67	.00
Error		129.12	160	.81			
Within-Subjects Effects							
Length	Linear	3.15	1	3.15	33.34	<.001**	.17
	Quadratic	1.12	1	1.12	16.77	<.001**	.10
Length × OCD-Group	Linear	.03	1	.03	.28	.60	.00
	Quadratic	.02	1	.02	.26	.61	.00
Length × Reasoning Condition	Linear	.26	1	.26	2.77	.10	.02
	Quadratic	.06	1	.06	.92	.34	.01
Length × OCD-Group × Reasoning Condition	Linear	.05	1	.05	.54	.46	.00
	Quadratic	.00	1	.00	.04	.85	.00
Error (Length)	Linear	15.10	160	.09			
	Quadratic	10.68	160	.07			

Note. **p* < .05, ***p* < .01.

Table A 1.6

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Confidence Ratings to Invalid Arguments

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Group		4.01	1	4.01	2.42	.12	.02
Reasoning Condition		.01	1	.01	.01	.94	.00
OCD-Group × Reasoning Condition		.00	1	.00	.00	.99	.00
Error		265.17	160	1.66			
Within-Subjects Effects							
Length	Linear	4.30	1	4.30	20.09	<.001**	.11
	Quadratic	.26	1	.26	3.32	.07	.02
Length × OCD-Group	Linear	.07	1	.07	.34	.56	.00
	Quadratic	.01	1	.01	.08	.78	.00
Length × Reasoning Condition	Linear	.52	1	.52	2.41	.12	.02
	Quadratic	.06	1	.06	.76	.38	.01
Length × OCD-Group × Reasoning Condition	Linear	.22	1	.22	1.04	.31	.01
	Quadratic	.11	1	.11	1.36	.25	.01
Error (Length)	Linear	32.23	160	.21			
	Quadratic	12.36	160	.08			

Note. **p* < .05, ***p* < .01.

Table A 1.7

Experiment 1 Premise Monotonicity Induction-Deduction Task Mean Proportion of Positive Arguments Accepted ('Strong' / 'Valid') for OCD Quartile Groups (Standard Deviations in Brackets)

		Lowest OCD quartile		Highest OCD Quartile	
Argument validity	Argument length	Deduction	Induction	Deduction	Induction
Valid	One-premise	1.00 (0.22)	0.96 (0.08)	0.87 (0.22)	0.85 (0.26)
	Three-premise	0.90 (0.13)	0.86 (0.20)	0.75 (0.19)	0.70 (0.30)
	Five-premise	0.87 (0.14)	0.83 (0.24)	0.74 (0.20)	0.75 (0.27)
Invalid	One-premise	0.02 (0.09)	0.05 (0.10)	0.15 (0.21)	0.14 (0.24)
	Three-premise	0.03 (0.07)	0.13 (0.24)	0.20 (0.25)	0.21 (0.29)
	Five-premise	0.09 (0.18)	0.17 (0.23)	0.23 (0.31)	0.33 (0.32)

Table A 1.8

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Proportion of Positive Arguments Accepted ('Strong'/ 'Valid') for OCD Quartile Groups

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Quartile		.00	1	.00	.00	.97	.00
Reasoning Condition		.02	1	.02	.17	.69	.00
OCD-Quartile × Reasoning Condition		.00	1	.00	.02	.89	.00
Error		9.06	82	.11			
Within-Subjects Effects							
Validity		62.01	1	62.01	629.92	<.001**	.89
Validity × OCD-Quartile		2.01	1	2.01	20.42	<.001**	.20
Validity × Reasoning Condition		.21	1	.21	2.11	.15	.03
Validity × OCD-Quartile × Reasoning Condition		.02	1	.02	.19	.67	.00
Error (Validity)		8.07	82	.10			
Length	Linear	.00	1	.00	.030	.86	.00
	Quadratic	.10	1	.10	8.70	.004**	.10
Length × OCD-Quartile	Linear	.02	1	.02	.49	.48	.01
	Quadratic	.01	1	.01	1.12	.29	.01
Length × Reasoning Condition	Linear	.05	1	.05	.150	.23	.02
	Quadratic	.00	1	.00	.26	.61	.00
Length × OCD-Quartile × Reasoning Condition	Linear	.01	1	.01	.22	.64	.00
	Quadratic	.02	1	.02	1.50	.23	.02
Error (Length)	Linear	2.95	82	.04			
	Quadratic	.94	82	.01			
Validity × Length	Linear	1.21	1	1.21	64.82	<.001**	.44
	Quadratic	.07	1	.07	6.41	.01*	.07
Error (Validity × Length)	Linear	1.53	82	.02			
	Quadratic	.95	82	.01			

Note. $*p < .05$, $**p < .01$.

Table A 1.9

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Proportion of Positive Responses (‘Strong’/ ‘Valid’) to Valid Items for OCD-Quartile Groups

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Quartile		.99	1	.99	12.50	.001**	.13
Reasoning Condition		.05	1	.05	.65	.42	.01
OCD-Quartile × Reasoning Condition		.00	1	.00	.05	.82	.00
Error		6.45	82	.08			
Within-Subjects Effects							
Length	Linear	.64	1	.64	19.32	<.001**	.19
	Quadratic	.17	1	.17	12.03	.001**	.13
Length × OCD-Quartile	Linear	.00	1	.00	.05	.83	.00
	Quadratic	.02	1	.02	.27	.61	.00
Length × Reasoning Condition	Linear	.00	1	.00	.03	.87	.00
	Quadratic	.01	1	.01	.62	.44	.01
Length × OCD-Quartile × Reasoning Condition	Linear	.00	1	.00	.05	.83	.00
	Quadratic	.00	1	.00	.27	.61	.00
Error (Length)	Linear	2.73	82	.03			
	Quadratic	1.18	82	.01			

Note. *$p < .05$, **$p < .01$.

Table A 1.9

Experiment 1 Premise Monotonicity Induction-Deduction Task ANOVA Summary Table for Proportion of Positive Responses (‘Strong’/ ‘Valid’) to Invalid Items for OCD-Quartile Groups

Effect	Trend	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects							
OCD-Quartile		1.02	1	1.02	7.87	.006**	.09
Reasoning Condition		.18	1	.18	1.35	.25	.02
OCD- Quartile × Reasoning Condition		.02	1	.02	.13	.72	.00
Error		10.64	82	.13			
Within-Subjects Effects							
Length	Linear	.57	1	.57	26.61	<.001**	.25
	Quadratic	.00	1	.00	.11	.74	.00
Length × OCD-Quartile	Linear	.02	1	.02	.92	.34	.01
	Quadratic	.00	1	.00	.01	.92	.00
Length × Reasoning Condition	Linear	.08	1	.09	4.12	.05*	.05
	Quadratic	.00	1	.00	.03	.86	.00
Length × OCD-Quartile × Reasoning Condition	Linear	.01	1	.01	.35	.56	.00
	Quadratic	.02	1	.02	1.76	.19	.02
Error (Length)	Linear	1.76	82	.02			
	Quadratic	.71	82	.01			

Note. *$p < .05$, **$p < .01$.

Appendix B

Chapter 3: Sensitivity to Premise Diversity and Argument Validity in Obsessive-Compulsive Disorder

Experiment 2: Item Set

Table B 1.1

Experiment 2 Premise Diversity Induction-Deduction Task Item Set

Item Content	Common Premise	Diverse Premise/Conclusion	Non-Diverse Premise/Conclusion	Superordinate Premise/Conclusion	Property
OCD-neutral	chimps	mice	gorillas	mammals	have a (ileal, azygos, cephalic, basilica) vein
	lions	bats	cats	mammals	use (dopamine, tyramine, serotonin, vasopressin) as a neurotransmitter
	wolves	pigs	foxes	mammals	require vitamin (B12, K, B6, E) for the liver to function
OCD-relevant	silver coins	dollar bills	gold coins	forms of money	contamined by the microbe (clostrisia mestoucis, chlosophilius pylori, plasmodium vasiporans, hemonasella coli)
	soiled shorts	soiled jumpers	soiled trousers	soiled clothing	can cause the disease (mylecoccus, cytomyelisis, respherocytis, bastosiridium)
	liquid soap	baby wipes	bars of soap	cleaning products	can remove encrusted (dirt, sediment, grime, filth)
	switched on desk lamps	switched on lanterns	switched on bedside lamps	switched on light sources	have a chance of producing a (hazardous, dangerous, harmful, perilous) spark
	plugged in microwaves	plugged in alarm clocks	plugged in toasters	plugged in electrical appliances	have a chance of causing (powerful, intense, strong, severe) shocks
	lit ovens	lit candles	lit stoves	lit heat producing devices	have a chance of causing a (wildfire, inferno, blaze, firestorm)
Attention check	dogs	birds	-	alcohol	require biotin for haemoglobin synthesis
	charcoal barbeques	gas barbeques	-	flowers	have a chance of combusting
	raw chicken	raw pork	-	furniture	contaminated by the microorganism trephyticus aureus

Note. Premises and the conclusion were always preceded by the word ‘all’, and followed by a randomly assigned property. Each argument contained the ‘common premise’. Valid arguments included the superordinate category as a premise and either the diverse or non-diverse option as a conclusion. Invalid arguments included the superordinate category as a conclusion, and the diverse or non-diverse option as a premise.

Experiment 2: Statistical Summaries

Table B 1.2

Experiment 2 Premise Diversity Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive Responses ('Strong'/ 'Valid')

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	.00	1	.00	.04	.84	.00
Reasoning Condition	1.76	1	1.76	14.14	<.001**	.11
OCD-Group × Reasoning Condition	.05	1	.05	.38	.54	.00
Error	13.83	111	.12			
Within-Subjects Effects						
OCD-Relevance	.73	1	.73	11.37	.001**	.09
OCD-Relevance × OCD-Group	.01	1	.01	.09	.77	.00
OCD-Relevance × Reasoning Condition	.12	1	.12	1.93	.17	.02
OCD-Relevance × OCD-Group × Reasoning Condition	.00	1	.00	.03	.86	.00
Error (OCD-Relevance)	7.24	112	.06			
Validity	82.76	1	82.76	359.88	<.001**	.76
Validity × OCD-Group	.80	1	.80	3.47	.07	.03
Validity × Reasoning Condition	4.91	1	4.91	21.36	<.001**	.16
Validity × OCD-Group × Reasoning Condition	.08	1	.08	.37	.55	.00
Error (Validity)	25.76	112	.23			
Diversity	.02	1	.02	.66	.42	.01
Diversity × OCD-Group	.06	1	.06	1.92	.17	.02
Diversity × Reasoning Condition	.04	1	.04	1.27	.26	.01
Diversity × OCD-Group × Reasoning Condition	.05	1	.05	1.45	.23	.01
Error (Diversity)	3.70	112	.03			
OCD-Relevance × Validity	.06	1	.06	1.38	.24	.01
OCD-Relevance × Validity × OCD-Group	.03	1	.03	.65	.42	.00
OCD-Relevance × Validity × Reasoning Condition	.08	1	.08	1.82	.18	.02
OCD-Relevance × Validity × OCD-Group × Reasoning Condition	.00	1	.00	.01	.91	.00
Error (OCD-Relevance × Validity)	5.08	112	.05			

Note. *p < .05, **p < .01.

Table B 1.2
Continued

Effect	SS	df	Mean Square	F	Sig	η_p^2
Within-Subjects Effects						
OCD-Relevance × Diversity	.01	1	.01	.49	.48	.00
OCD-Relevance × Diversity × OCD-Group	.00	1	.00	.04	.85	.00
OCD-Relevance × Diversity × Reasoning Condition	.02	1	.02	.67	.42	.01
OCD-Relevance × Diversity × OCD-Group × Reasoning Condition	.00	1	.00	.00	.98	.00
Error (OCD-Relevance × Diversity)	3.06	112	.03			
Validity × Diversity	.79	1	.79	20.78	<.001**	.16
Validity × Diversity × OCD-Group	.00	1	.00	.10	.75	.00
Validity × Diversity × Reasoning Condition	.36	1	.36	9.55	.003**	.08
Validity × Diversity × OCD-Group × Reasoning Condition	.00	1	.00	.00	.99	.00
Error (Validity × Diversity)	4.27	112	.04			
OCD-Relevance × Validity × Diversity	.01	1	.01	.28	.60	.00
OCD-Relevance × Validity × Diversity × OCD-Group	.02	1	.02	1.10	.30	.01
OCD-Relevance × Validity × Diversity × Reasoning Condition	.00	1	.00	.25	.62	.00
OCD-Relevance × Validity × Diversity × OCD-Group × Reasoning Condition	.00	1	.00	.04	.83	.00
Error (OCD-Relevance × Validity × Diversity)	2.07	112	.02			

Note. **p* < .05, ***p* < .01.

Table B 1.3

Experiment 2 Premise Diversity Induction-Deduction Task ANCOVA Summary Table for Confidence Ratings by OCD-Group and Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	5.48	1	5.48	2.67	.10	.02
Reasoning Condition	13.72	1	13.72	6.74	.01*	.06
OCD-Group × Reasoning Condition	1.70	1	1.70	.84	.36	.01
Error	226.02	111	2.04			
Within-Subjects Effects						
OCD-Relevance	.79	1	.79	2.21	.14	.02
OCD-Relevance × OCD-Group	.00	1	.00	.00	1.00	.00
OCD-Relevance × Reasoning Condition	.28	1	.28	.79	.38	.01
OCD-Relevance × OCD-Group × Reasoning Condition	.88	1	.88	2.45	.12	.02
Error (OCD-Relevance)	40.14	112	.36			
Validity	22.34	1	22.34	29.98	<.001**	.21
Validity × OCD-Group	1.14	1	1.14	1.53	.22	.01
Validity × Reasoning Condition	3.74	1	3.74	5.02	.03*	.04
Validity × OCD-Group × Reasoning Condition	1.92	1	1.92	2.58	.11	.02
Error (Validity)	83.43	112	.74			
Diversity	1.94	1	1.94	10.83	.001**	.09
Diversity × OCD-Group	.00	1	.00	.00	1.00	.00
Diversity × Reasoning Condition	.08	1	.08	.47	.49	.00
Diversity × OCD-Group × Reasoning Condition	.43	1	.43	2.40	.12	.02
Error (Diversity)	20.03	112	.18			
OCD-Relevance × Validity	1.24	1	1.24	6.92	.01*	.06
OCD-Relevance × Validity × OCD-Group	.23	1	.23	1.27	.26	.01
OCD-Relevance × Validity × Reasoning Condition	.01	1	.01	.05	.83	.00
OCD-Relevance × Validity × OCD-Group × Reasoning Condition	.08	1	.08	.46	.50	.00
Error (OCD-Relevance × Validity)	20.07	112	.18			

Note. *$p < .05$, **$p < .01$.

Table B 1.3

Continued

Effect	SS	df	Mean Square	F	Sig	η_p^2
Within-Subjects Effects						
OCD-Relevance × Diversity	.90	1	.90	6.34	.01*	.05
OCD-Relevance × Diversity × OCD-Group	.00	1	.00	.01	.91	.00
OCD-Relevance × Diversity × Reasoning Condition	.00	1	.00	.01	.91	.00
OCD-Relevance × Diversity × OCD-Group × Reasoning Condition	.01	1	.01	.06	.80	.00
Error (OCD-Relevance × Diversity)	15.87	112	.14			
Validity × Diversity	2.48	1	2.48	18.37	<.001**	.14
Validity × Diversity × OCD-Group	.05	1	.05	.34	.56	.00
Validity × Diversity × Reasoning Condition	.34	1	.34	2.51	.12	.02
Validity × Diversity × OCD-Group × Reasoning Condition	.47	1	.47	3.48	.07	.03
Error (Validity × Diversity)	15.14	112	.14			
OCD-Relevance × Validity × Diversity	.01	1	.01	.07	.80	.00
OCD-Relevance × Validity × Diversity × OCD-Group	.02	1	.02	.19	.66	.00
OCD-Relevance × Validity × Diversity × Reasoning Condition	.13	1	.13	1.09	.30	.01
OCD-Relevance × Validity × Diversity × OCD-Group × Reasoning Condition	.03	1	.03	.24	.62	.00
Error (OCD-Relevance × Validity × Diversity)	12.97	112	.12			

Note. $*p < .05$, $**p < .01$.

Table B 1.4

Experiment 2 Premise Diversity Induction-Deduction Task Mean Proportion of Positive ('Strong'/ 'Valid') Responses by OCD-Quartile

OCD-relevance	Argument type	Lowest OCD quartile		Highest OCD quartile	
		Deduction	Induction	Deduction	Induction
Neutral	Valid/Diverse	0.82 (0.25)	0.88 (0.26)	0.79 (0.36)	0.80 (0.29)
	Valid/Non-Diverse	0.85 (0.29)	0.96 (0.16)	0.85 (0.30)	0.93 (0.14)
	Invalid/Diverse	0.07 (0.26)	0.33 (0.37)	0.19 (0.37)	0.33 (0.39)
	Invalid/Non-Diverse	0.07 (0.19)	0.18 (0.36)	0.21 (0.34)	0.26 (0.24)
Relevant	Valid/Diverse	0.93 (0.14)	0.91 (0.12)	0.83 (0.23)	0.80 (0.33)
	Valid/Non-Diverse	0.96 (0.08)	0.92 (0.12)	0.88 (0.25)	0.92 (0.11)
	Invalid/Diverse	0.06 (0.17)	0.34 (0.30)	0.24 (0.37)	0.41 (0.37)
	Invalid/Non-Diverse	0.07 (0.15)	0.20 (0.29)	0.26 (0.39)	0.32 (0.34)

Table B 1.5

Experiment 2 Premise Diversity Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive ('Strong'/ 'Valid') Responses by OCD-Quartile and Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Quartile	.04	1	.04	.31	.58	.01
Reasoning Condition	.74	1	.74	5.62	.02*	.09
OCD-Quartile × Reasoning Condition	.07	1	.07	.56	.46	.01
Error	7.38	56	.13			
Within-Subjects Effects						
OCD-Relevance	.13	1	.13	2.30	.14	.04
OCD-Relevance × OCD-Quartile	.03	1	.03	.49	.49	.01
OCD-Relevance × Reasoning Condition	.00	1	.00	.02	.88	.00
OCD-Relevance × OCD-Quartile × Reasoning Condition	.01	1	.01	.20	.66	.00
Error (OCD-Relevance)	3.23	57	.06			
Validity	52.15	1	52.15	211.20	<.001**	.79
Validity × OCD-Quartile	1.09	1	1.09	4.41	.04*	.07
Validity × Reasoning Condition	.66	1	.66	2.67	.11	.04
Validity × OCD-Quartile × Reasoning Condition	.06	1	.06	.22	.64	.00
Error (Validity)	14.08	57	.25			

Note. *p < .05, **p < .01.

Table B 1.5
Continued

Effect	SS	df	Mean Square	F	Sig	η_p^2
Within-Subjects Effects						
Diversity	.00	1	.00	.09	.76	.00
Diversity × OCD-Quartile	.02	1	.02	.65	.42	.01
Diversity × Reasoning Condition	.11	1	.11	3.01	.09	.05
Diversity × OCD-Quartile × Reasoning Condition	.02	1	.02	.64	.43	.01
Error (Diversity)	2.17	57	.04			
OCD-Relevance × Validity	.00	1	.00	.00	1.00	.00
OCD-Relevance × Validity × OCD-Quartile	.02	1	.02	.52	.47	.01
OCD-Relevance × Validity × Reasoning Condition	.02	1	.02	.50	.48	.01
OCD-Relevance × Validity × OCD-Quartile × Reasoning Condition	.01	1	.01	.31	.58	.01
Error (OCD-Relevance × Validity)	2.29	57	.04			
OCD-Relevance × Diversity	.00	1	.00	.15	.70	.00
OCD-Relevance × Diversity × OCD-Quartile	.02	1	.02	.54	.47	.01
OCD-Relevance × Diversity × Reasoning Condition	.01	1	.01	.18	.67	.00
OCD-Relevance × Diversity × OCD-Quartile × Reasoning Condition	.00	1	.00	.07	.79	.00
Error (OCD-Relevance × Diversity)	1.61	57	.03			
Validity × Diversity	.40	1	.40	8.44	.005**	.13
Validity × Diversity × OCD-Quartile	.01	1	.01	.11	.74	.00
Validity × Diversity × Reasoning Condition	.14	1	.14	3.04	.09	.05
Validity × Diversity × OCD-Quartile × Reasoning Condition	.00	1	.00	.00	.97	.00
Error (Validity × Diversity)	2.71	57	.05			
OCD-Relevance × Validity × Diversity	.01	1	.01	.28	.60	.00
OCD-Relevance × Validity × Diversity × OCD-Quartile	.03	1	.03	1.41	.24	.02
OCD-Relevance × Validity × Diversity × Reasoning Condition	.01	1	.01	.40	.53	.01
OCD-Relevance × Validity × Diversity × OCD-Quartile × Reasoning Condition	.00	1	.00	.18	.67	.00
Error (OCD-Relevance × Validity × Diversity)	1.17	57	.02			

Note. $*p < .05$, $**p < .01$.

Experiment 3: Item Set

Table B1.6

Experiment 3 Forced-Choice Premise-Diversity Task Item Set

Common Premise	Diverse Premise	Non-Diverse Premise	Conclusion	Property
OCD-neutral items				
chimps	mice	gorillas	mammals	have a ileal vein
hippos	rats	elephants	mammals	require trace amounts of magnesium for reproduction
cows	bats	zebras	mammals	require biotin for haemoglobin synthesis
elephants	gorillas	rhinos	mammals	have a higher potassium concentration in their blood than humans
camels	squirrels	horses	mammals	have lutein in their cells
goats	possums	donkeys	mammals	secrete uric acid crystals
rhinos	sheep	hippos	mammals	have sesamoid bones
moose	rabbits	deer	mammals	have a higher sodium concentration in their blood than humans
foxes	seals	dogs	mammals	have a choroid membrane in their eyes
deer	dolphins	donkeys	mammals	have a lower body temperature at infancy than at maturity
wolves	pigs	foxes	mammals	require Vitamin K for the liver to function
lions	bats	cats	mammals	use serotonin as a neurotransmitter
squirrels	zebras	rats	mammals	require vitamin (B12, K, B6, E) for the liver to function
horses	mice	cows	mammals	use the neurotransmitter Dihedron
dolphins	chimps	seals	mammals	require oxydilic acid for good digestion
OCD-relevant items				
silver coins	dollar bills	gold coins	forms of money	are contamined by the germ hemonasella coli
soiled shorts	soiled jumpers	soiled trousers	soiled clothing	can cause the disease respherocytis
liquid soap	baby wipes	bars of soap	cleaning products	can remove stains that are protein based
used toilet paper	used writing paper	used tissue paper	types of used paper	contain the chemical didymium arsenide
sweaty shirts	sweaty underpants	sweaty singlets	sweaty clothing	contain the yeast pacyliococcus maliformans
raw chicken	raw pork	raw turkey	raw meat	is contaminated by the microorganism chlosophilius pylori
switched on desk lamps	switched on lanterns	switched on bedside lamps	switched on light sources	have a chance of producing a dangerous spark
plugged in microwaves	plugged in alarm clocks	plugged in toasters	plugged in electrical appliances	have a chance of causing electrical shocks
lit ovens	lit candles	lit stoves	lit heat producing devices	have a chance of causing an inferno
front doors	car doors	screen doors	doors	have a chance of failing to lock
exams	essays	quizzes	types of assessment	have a chance of containing critical errors
meat cleavers	butter knives	carving knives	knives	can damage corsucules under the skin
drills	handsaws	screwdrivers	tools	can damage the epidermis
used hypodermic needles	used sewing needles	used suture needles	used needles	can cause a puncture wound
broken kitchen windows	broken car windows	broken bathroom windows	broken windows	can cause a haemorrhage

Note. Each premise and conclusion was preceded by "All" and followed by the property.

Table B1.7

Experiment 3 and 4 Similarity Ratings for Diverse and Non-Diverse Item Pairs

Common Premise	Diverse Premise	Non-Diverse Premise	Conclusion	Similarity of category to conclusion	
				Diverse	Non-diverse
OCD-neutral items					
chimps	mice	gorillas	mammals	3.79 (1.81)	9.00 (1.11)
hippos	rats	elephants	mammals	4.05 (1.93)	7.63 (1.38)
cows	bats	zebras	mammals	3.53 (1.26)	6.90 (1.24)
elephants	gorillas	rhinos	mammals	4.32 (1.57)	7.53 (1.35)
camels	squirrels	horses	mammals	4.16 (2.17)	7.42 (1.30)
goats	possums	donkeys	mammals	4.32 (2.36)	7.47 (1.17)
rhinos	sheep	hippos	mammals	4.74 (2.10)	7.90 (1.10)
moose	rabbits	deer	mammals	4.79 (1.81)	7.37 (1.54)
foxes	seals	dogs	mammals	3.74 (2.05)	7.37 (1.42)
deer	dolphins	donkeys	mammals	3.90 (1.97)	7.79 (1.32)
wolves	pigs	foxes	mammals	3.90 (1.70)	7.79 (1.55)
lions	bats	cats	mammals	4.05 (2.20)	8.32 (1.49)
squirrels	zebras	rats	mammals	4.00 (1.63)	7.26 (1.41)
horses	mice	cows	mammals	4.00 (2.29)	7.63 (1.26)
dolphins	chimps	seals	mammals	3.79 (2.10)	7.63 (1.07)
OCD-relevant items					
silver coins	dollar bills	gold coins	forms of money	4.53 (2.12)	8.63 (1.30)
soiled shorts	soiled jumpers	soiled trousers	soiled clothing	4.90 (2.08)	8.21 (1.27)
liquid soap	baby wipes	bars of soap	cleaning products	5.11 (2.28)	8.26 (1.33)
used toilet paper	used writing paper	used tissue paper	types of used paper	4.84 (2.41)	7.37 (1.61)
sweaty shirts	sweaty underpants	sweaty singlets	sweaty clothing	5.63 (2.27)	8.05 (1.31)
raw chicken	raw pork	raw turkey	raw meat	5.53 (2.14)	8.63 (1.34)
switched on desk lamps	switched on lanterns	switched on bedside lamps	switched on light sources	4.26 (2.05)	8.47 (1.12)
plugged in microwaves	plugged in alarm clocks	plugged in toasters	plugged in electrical appliances	5.00 (2.7)	8.16 (1.21)
lit ovens	lit candles	lit stoves	lit heat producing devices	5.32 (2.03)	7.90 (2.00)
front doors	car doors	screen doors	doors	5.32 (2.60)	7.74 (1.52)
exams	essays	quizzes	types of assessment	5.21 (1.99)	7.74 (1.37)
meat cleavers	butter knives	carving knives	knives	5.00 (1.91)	7.68 (1.73)
drills	handsaws	screwdrivers	tools	5.84 (2.09)	7.68 (1.25)
used hypodermic needles	used sewing needles	used suture needles	used needles	6.53 (2.74)	7.63 (2.27)
broken kitchen windows	broken car windows	broken bathroom windows	broken windows	5.90 (2.05)	7.42 (1.74)

Experiment 3: Statistical Summaries

Table B 1.8

Experiment 3 Forced-Choice Diversity Task ANCOVA Summary Table for Proportion of Diverse Arguments Endorsed, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	.34	1	.34	2.69	.10	.03
Error	12.41	97	.13			
Within-Subjects Effects						
OCD-Relevance	.00	1	.00	.04	.85	.00
OCD-Relevance × OCD-Group	.01	1	.01	.56	.46	.01
Error (OCD-Relevance)	1.31	98	.01			

Note. **p* < .05, ***p* < .01.

Table B 1.9

Experiment 3 Forced-Choice Diversity Task ANCOVA Summary Table for Perceived Usefulness Ratings, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	2.44	1	2.44	.29	.59	.00
Error	806.51	97	8.31			
Within-Subjects Effects						
OCD-Relevance	.35	1	.35	1.27	.26	.01
OCD-Relevance × OCD-Group	.94	1	.94	3.40	.07	.03
Error (OCD-Relevance)	27.05	98	.28			
Diversity	78.43	1	78.43	34.05	<.001**	.26
Diversity × OCD-Group	5.35	1	5.35	2.32	.13	.02
Error (Diversity)	225.72	98	2.30			
OCD-Relevance × Diversity	.00	1	.00	.00	.98	.00
OCD-Relevance × Diversity × OCD-Group	.18	1	.18	1.10	.30	.01
Error (Content × Diversity)	16.47	98	.17			

Note. **p* < .05, ***p* < .01.

Table B 1.10

Experiment 3 Forced-Choice Diversity Task ANCOVA Summary Table for Proportion of Diverse Arguments Endorsed by OCD-Quartile, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Quartile	.96	1	.96	10.09	.003**	.17
Error	4.59	48	.10			
Within-Subjects Effects						
OCD-Relevance	.00	1	.00	.12	.74	.00
OCD-Relevance × OCD-Quartile	.02	1	.02	2.04	.16	.04
Error (OCD-Relevance)	.52	49	.01			

Note. *p < .05, **p < .01.

Table B 1.11

Experiment 3 Forced-Choice Diversity Task ANCOVA Summary Table for Perceived Usefulness Ratings by OCD-Quartile, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Quartile	.00	1	.00	.00	.99	.00
Error	356.36	48	7.42			
Within-Subjects Effects						
OCD-Relevance	.01	1	.01	.03	.86	.00
OCD-Relevance × OCD-Quartile	.86	1	.86	3.28	.08	.06
Error (OCD-Relevance)	12.87	49	.26			
Diversity	45.24	1	45.24	28.05	<.001**	.36
Diversity × OCD-Quartile	14.14	1	14.14	8.76	.005**	.15
Error (Diversity)	79.03	49	1.61			
OCD-Relevance × Diversity	.00	1	.00	.01	.92	.00
OCD-Relevance × Diversity × OCD-Quartile	.25	1	.25	3.09	.09	.06
Error (Content × Diversity)	3.90	49	.08			

Note. *p < .05, **p < .01.

Experiment 4: Item Set

Table B1.12

Experiment 4 Hypothesis Testing Premise-Diversity Task Item Set

Given premise	Diverse premises	Intermediately diverse premises	Non-diverse premises	Conclusion	Property
Neutral					
chimps	seals porpoises	mice hamsters	gorillas orangutans	mammals	have a choroid membrane in their eyes
cats	bats beavers	wolves foxes	lions tigers	mammals	have sesamoid bones
hawks	penguins emus	crows magpies	eagles falcons	birds	have an ileal vein
oranges	watermelons coconuts	pears apples	lemons mandarins	fruits	produce the germination agent coumarin
OCD-relevant					
gold coins	credit cards debit cards	10 dollar bills 20 dollar bills	silver coins bronze coins	forms of money	are contaminated by the germ hemonasella coli
soiled trousers	soiled coats soiled jumpers	soiled dresses soiled skirts	soiled jeans soiled sweatpants	soiled clothing	can cause the disease respherocytis
dirty tablespoons	dirty plates dirty bowls	dirty knives dirty forks	dirty soup spoons dirty dessert spoons	dirty dinnerware	has traces of the microbe plasmodium vasiporans
raw turkey	raw crab raw fish	raw lamb raw pork	raw duck raw chicken	forms of raw meat	are contaminated by the microorganism chlosophilius pylori

Experiment 4: Statistical Summaries

Table B1.13

Experiment 4 Hypothesis Testing Premise-Diversity Task ANCOVA Summary Table for Critical Variables of Interest, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Overall number tested						
OCD-Group	.10	1	.10	.05	.83	.00
Error	325.96	159	2.05			
OCD-Relevance	.05	1	.05	.16	.69	.00
OCD-Group × OCD-Relevance	.02	1	.02	.06	.81	.00
Error (OCD-Relevance)	50.02	160	.31			
Overall diverse						
OCD-Group	.09	1	.09	1.81	.18	.01
Error	7.71	159	.05			
OCD-Relevance	.56	1	.56	50.94	<.001**	.24
OCD-Group × OCD-Relevance	.01	1	.01	.46	.50	.00
Error (OCD-Relevance)	1.76	160	.01			
Overall non-diverse						
OCD-Group	.10	1	.10	1.58	.21	.01
Error	10.37	159	.07			
OCD-Relevance	.09	1	.09	9.39	.003**	.06
OCD-Group × OCD-Relevance	.01	1	.01	1.23	.27	.01
Error (OCD-Relevance)	1.61	160	.01			
Diverse tested first						
OCD-Group	.02	1	.02	.08	.77	.00
Error	30.98	159	.19			
OCD-Relevance	.06	1	.06	1.84	.18	.01
OCD-Group × OCD-Relevance	.01	1	.01	.14	.71	.00
Error (OCD-Relevance)	5.62	160	.04			
Non-diverse tested first						
OCD-Group	.08	1	.08	.42	.52	.00
Error	31.68	159	.20			
OCD-Relevance	.12	1	.12	3.43	.07	.02
OCD-Group × OCD-Relevance	.01	1	.01	.23	.63	.00
Error (OCD-Relevance)	5.78	160	.04			
Non-diverse perseveration						
OCD-Group	.15	1	.15	1.12	.29	.01
Error	21.53	159	.14			
OCD-Relevance	.02	1	.02	.72	.40	.00
OCD-Group × OCD-Relevance	.00	1	.00	.00	.98	.00
Error (OCD-Relevance)	3.48	160	.02			

Note. **p* < .05, ***p* < .01.

Table B1.14

Experiment 4 Hypothesis Testing Premise-Diversity Task Means for Critical Variables of Interest by OCD-Quartile (Standard Deviations in Parentheses)

	OCD-neutral items		OCD-relevant items	
Variable of interest	Low OCD quartile	High OCD quartile	Low OCD quartile	High OCD quartile
Overall number tested	3.28 (1.14)	3.42 (1.10)	3.30 (1.13)	3.49 (1.15)
Proportion diverse	0.49 (0.20)	0.43 (0.22)	0.39 (0.12)	0.35 (0.15)
Proportion non-diverse	0.19 (0.20)	0.24 (0.23)	0.23 (0.17)	0.29 (0.18)
Proportion diverse first	0.47 (0.38)	0.42 (0.38)	0.44 (0.32)	0.42 (0.33)
Proportion non-diverse first	0.29 (0.34)	0.31 (0.38)	0.34 (0.34)	0.35 (0.35)
Non-diverse perseveration	0.19 (0.27)	0.20 (0.34)	0.15 (0.24)	0.22 (0.30)

Table B1.15

Experiment 4 Hypothesis Testing Premise-Diversity Task ANCOVA Summary Table for Critical Variables of Interest by OCD-Quartile, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Overall number tested						
OCD-Quartile	1.03	1	1.03	.43	.51	.01
Error	196.00	82	2.39			
OCD-Relevance	.07	1	.07	.38	.54	.00
OCD-Quartile × OCD-Relevance	.02	1	.02	.12	.73	.00
Error (OCD-Relevance)	15.72	83	.19			
Overall diverse						
OCD-Quartile	.08	1	.08	1.61	.21	.02
Error	4.24	82	.05			
OCD-Relevance	.34	1	.34	35.26	<.001**	.30
OCD-Quartile × OCD-Relevance	.00	1	.00	.38	.54	.00
Error (OCD-Relevance)	.80	83	.01			
Overall non-diverse						
OCD-Quartile	.12	1	.12	1.79	.19	.02
Error	5.65	82	.07			
OCD-Relevance	.10	1	.10	14.53	<.001**	.15
OCD-Quartile × OCD-Relevance	.00	1	.00	.02	.90	.00
Error (OCD-Relevance)	.58	83	.01			
Diverse tested first						
OCD-Quartile	.04	1	.04	.20	.65	.00
Error	17.74	82	.22			
OCD-Relevance	.01	1	.01	.31	.58	.00
OCD-Quartile × OCD-Relevance	.01	1	.01	.31	.58	.00
Error (OCD-Relevance)	2.51	83	.03			
Non-diverse tested first						
OCD-Quartile	.02	1	.02	.08	.78	.00
Error	17.24	82	.21			
OCD-Relevance	.08	1	.08	2.55	.11	.03
OCD-Quartile × OCD-Relevance	.00	1	.00	.02	.90	.00
Error (OCD-Relevance)	2.70	83	.03			
Non-diverse perseveration						
OCD-Quartile	.06	1	.06	.44	.51	.01
Error	11.55	82	.14			
OCD-Relevance	.00	1	.00	.19	.66	.00
OCD-Quartile × OCD-Relevance	.02	1	.02	1.20	.28	.01
Error (OCD-Relevance)	1.51	83	.02			

Note. $^*p < .05$, $^{**}p < .01$.

Experiment 5: Statistical Summaries

Table B1.16

Experiment 5 Similarity Rating Task ANCOVA Summary Table for Proportion of Matching Responses Given, Controlling for Proxy General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	.21	1	.21	18.99	<.001**	.07
Error	2.92	258	.01			
Within-Subjects Effects						
OCD-Relevance	.66	1	.66	217.51	<.001**	.46
OCD-Relevance × OCD-Group	.03	1	.03	11.17	.001**	.04
Error (OCD-Relevance)	.78	259	.00			

Note. *p < .05, **p < .01.

Table B1.17

Experiment 5 Similarity Rating Task ANCOVA Summary Table for Similarity Ratings, Controlling for Proxy General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	1.28	1	1.28	.27	.60	.00
Error	1205.16	258	4.67			
Within-Subjects Effects						
OCD-Relevance	172.38	1	172.38	283.79	<.001**	.52
OCD-Relevance × OCD-Group	.61	1	.61	1.01	.32	.00
Error (OCD-Relevance)	157.32	259	.61			
Diversity	2050.04	1	2050.04	671.89	<.001**	.72
Diversity × OCD-Group	.70	1	.70	.23	.63	.00
Error (Diversity)	790.25	259	3.05			
OCD-Relevance × Diversity	80.99	1	80.99	227.43	<.001**	.47
OCD-Relevance × Diversity × OCD-Group	.34	1	.34	.97	.33	.00
Error (OCD-Relevance × Diversity)	92.24	259	.36			

Note. *p < .05, **p < .01.

Appendix C

Chapter 4: Sensitivity to Background Knowledge and Argument Validity in Obsessive-Compulsive Disorder

Experiment 6: Item Set

Table C1.1

Experiment 6 Causal Consistency Induction-Deduction Task Item Set

First premise	Second premise	Conclusion
Modus ponens arguments		
If Jill rolls in the mud, Jill gets dirty (*clean*)	Jill rolls in the mud	Jill gets dirty (*clean*)
If Tom throws a wine glass, the wine glass breaks (*bounces*)	Tom throws a wine glass	The wine glass breaks (*bounces*)
If Mel drinks a glass of water, the glass empties (*fills*)	Mel drinks a glass of water	The glass empties (*fills*)
If John drops Ball 16, Ball 16 falls (*rises*)	John drops Ball 16	Ball 16 falls (*rises*)
If Fred sprinkles water on wood shavings, the shavings get wet (*ignite*)	Fred sprinkles water on wood shavings	The shavings get wet (*ignite*)
If Car #10 runs into a brick wall, it will stop (*speed up*)	Car #10 runs into a brick wall	Car #10 stops (*speeds up*)
If George pains the wall green, the wall turns green (*red*)	George paints the wall green	The wall turns green (*red*)
If Reggie locks the door, no one (*everyone*) can get in	Reggie locks the door	No one (*everyone*) can get in
Conjunctive syllogisms		
It is not true both that it is raining outside and it is dry (*wet*) outside	It is raining outside	It is not dry (*wet*) outside
It is not true both that the log is on fire and the log is cold (*hot*)	The log is on fire	The log is not cold (*hot*)
It is not true both that Michelle needs reading glasses and Michelle has perfect (*impaired*) vision	Michelle needs glasses	Michelle does not have perfect (*impaired*) vision
It is not true both that Alfred is walking on a beach and Alfred is walking on marble (*sand*)	Alfred is walking on a beach	Alfred is not walking on marble (*sand*)
It is not true both that Jenny is running and Jenny is sitting down (*Jenny is on her feet*)	Jenny is running	Jenny is not sitting down (*Jenny is not on her feet*)
It is not true both that the tree has been cut down and the tree starts (*stops*) growing	The tree has been cut down	The tree does not start (*stop*) growing
It is not true both that Tabitha is doing work and Tabitha is sleeping (*awake*)	Tabitha is doing work	Tabitha is not sleeping (*awake*)
It is not true both that the ice-cream is out in the sun and the ice-cream is beginning to freeze (*melt*)	The ice-cream is out in the sun	The ice-cream is not beginning to freeze (*melt*)

Note. Valid arguments contained both the first and second premises; invalid arguments only contained the second premise. Causally inconsistent content in italics and parentheses. Stimuli adapted from Rips (2001).

Experiment 6: Statistical Summaries

Table C1.2

Experiment 6 Causal Consistency Induction-Deduction Task ANOVA Summary Table for Proportion of Positive ('Strong'/ 'Valid') Responses

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	.00	1	.00	.07	.80	.00
Reasoning Condition	.00	1	.00	.05	.82	.00
OCD-Group × Reasoning Condition	.05	1	.05	.80	.37	.01
Error	6.14	.95	.06			
Within-Subjects Effects						
Validity	17.59	1	17.59	222.18	<.001**	.70
Validity × OCD-Group	.01	1	.01	.16	.69	.00
Validity × Reasoning Condition	.59	1	.59	7.43	.01**	.07
Validity × OCD-Group × Reasoning Condition	.13	1	.13	1.63	.21	.02
Error (Validity)	7.60	96	.08			
Consistency	14.72	1	14.72	255.89	<.001**	.73
Consistency × OCD-Group	.01	1	.01	.20	.65	.00
Consistency × Condition	.56	1	.56	9.80	.002**	.09
Consistency × OCD-Group × Reasoning Condition	.01	1	.01	.22	.64	.00
Error (Consistency)	5.52	96	.06			
Validity × Consistency	3.27	1	3.27	69.89	<.001**	.42
Validity × Consistency × OCD-Group	.02	1	.02	.39	.53	.00
Validity × Consistency × Reasoning Condition	.01	1	.01	.28	.60	.00
Validity × Consistency × OCD-Group × Reasoning Condition	.11	1	.11	2.35	.13	.02
Error (Validity × Consistency)	4.49	96	.05			

Note. *p < .05, **p < .01.

Table C1.3

Experiment 6 Causal Consistency Induction-Deduction Task ANOVA Summary Table for Confidence Ratings

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	1.66	1	1.66	1.57	.21	.02
Reasoning Condition	2.93	1	2.93	2.78	.10	.03
OCD-Group × Reasoning Condition	7.23	1	7.23	6.87	.01*	.07
Error	100.00	95	1.05			
Within-Subjects Effects						
Validity	4.76	1	4.76	14.99	<.001**	.14
Validity × OCD-Group	.14	1	.14	.45	.51	.00
Validity × Reasoning Condition	.03	1	.03	.09	.77	.00
Validity × OCD-Group × Reasoning Condition	.02	1	.02	.06	.80	.00
Error (Validity)	30.46	96	.32			
Consistency	2.59	1	2.59	9.91	.002**	.09
Consistency × OCD-Group	.03	1	.03	.11	.74	.00
Consistency × Reasoning Condition	.39	1	.39	1.50	.22	.02
Consistency × OCD-Group × Reasoning Condition	.01	1	.01	.03	.86	.00
Error (Consistency)	25.14	96	.26			
Validity × Consistency	2.13	1	2.13	9.59	.003**	.09
Validity × Consistency × OCD-Group	.03	1	.03	.14	.71	.00
Validity × Consistency × Reasoning Condition	.49	1	.49	2.21	.14	.02
Validity × Consistency × OCD-Group × Reasoning Condition	.39	1	.39	1.77	.19	.02
Error (Validity × Consistency)	21.33	96	.22			

Note. *p < .05, **p < .01.

Table C1.4

Experiment 6 Causal Consistency Induction-Deduction Task Mean Proportion of Positive ('Strong' / 'Valid') Responses by OCD-Quartile

	Lowest OCD symptom quartile		Highest OCD symptom quartile	
Argument type	Deduction	Induction	Deduction	Induction
Valid/Consistent	0.92 (0.16)	0.83 (0.20)	0.88 (0.13)	0.88 (0.16)
Valid/Inconsistent	0.79 (0.30)	0.67 (0.37)	0.73 (0.38)	0.55 (0.31)
Invalid/Consistent	0.60 (0.38)	0.81 (0.22)	0.55 (0.38)	0.75 (0.20)
Invalid/Inconsistent	0.08 (0.12)	0.02 (0.07)	0.02 (0.08)	0.09 (0.18)

Table C1.5

Experiment 6 Causal Consistency Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive ('Strong'/ 'Valid') Responses by OCD-Quartile

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Quartile	.09	1	.09	1.30	.26	.03
Reasoning Condition	.00	1	.00	.00	.99	.00
OCD-Quartile × Reasoning Condition	.01	1	.01	.14	.71	.00
Error	3.22	46	.07			
Within-Subjects Effects						
Validity	8.56	1	8.56	107.03	<.001**	.69
Validity × OCD-Quartile	.00	1	.00	.04	.84	.00
Validity × Reasoning Condition	.53	1	.53	6.59	.01*	.12
Validity × OCD-Quartile × Reasoning Condition	.01	1	.01	.13	.73	.00
Error (Validity)	3.76	47	.08			
Consistency	8.35	1	8.35	151.64	<.001**	.76
Consistency × OCD-Quartile	.00	1	.00	.05	.82	.00
Consistency × Reasoning Condition	.30	1	.30	5.36	.03*	.10
Consistency × OCD-Quartile × Reasoning Condition	.00	1	.00	.00	.97	.00
Error (Consistency)	2.59	47	.06			
Validity × Consistency	2.29	1	2.29	55.54	<.001**	.54
Validity × Consistency × OCD-Quartile	.08	1	.08	2.05	.16	.04
Validity × Consistency × Reasoning Condition	.03	1	.03	.71	.40	.01
Validity × Consistency × OCD-Quartile × Reasoning Condition	.05	1	.05	1.33	.26	.03
Error (Validity × Consistency)	1.94	47	.04			

Note. **p* < .05, ***p* < .01.

Experiment 7: Item Set

Table C1.6

Experiment 7 Belief Bias Task Item Set

Predicate (X) and middle (Z) terms		
OCD-neutral	OCD-relevant	Subject (Y)
animals - llamas	stained clothing - stained underwear	gebblers
birds - parrots	dirty cleaning products - dirty sponges	emblers
storms - blizzards	odorous public facilities - odorous toilets	glax
drinks - beers	used first aid products - used bandaids	sables
bears - grizzlies	mouldy oral hygiene products - mouldy toothbrush	krabbers
dances - tangoes	curdled dairy products - curdled milk	rabs
writers - novelists	saliva-covered utensils - saliva-covered chopsticks	brox
reptiles - lizards	sticky furniture - sticky tables	phylones
cars - sedans	used medical equipment - used suture needles	bictoids
killers - assassins	lit cooking equipment - lit gas BBQs	cryptods
trees - oaks	switched on lights - switched on lamps	zaphods
relatives - uncles	switched on kitchen appliances - switched on gas ovens	junarics
insects - grasshoppers	running water sources - running taps	renculions
criminals - robbers	plugged in appliances - plugged in toasters	glissomae
boats - canoes	broken glass types - broken glass bottles	hammerkops
words - verbs	sharp cutting devices - sharp carving knives	metazoans

Note. The assignment of premises to predicate or middle position depended on whether conclusions were believable or unbelievable. Content was randomised across problem structures (see Table C1.6) for each participant, and nonsensical subject premise terms were randomised across items.

Table C1.7

Experiment 7 Belief Bias Task Problem Structures

Set	First premise	Second premise	Conclusion
A	No X are Y	Some Z are Y	Some Z are not X
	Some X are Y	No Z are Y	Some X are not Z
B	Some X are not Y	All Z are Y	Some X are not Z
	No X are Y	Some Z are not Y	Some X are not Z

Note. Invalid items featured reversed ordering of conclusion terms (e.g. Some Z are not X becomes some X are not Z). Believability was manipulated by reversing the order of premise content. Set A includes problems that minimise figure, atmosphere and conversion effects. Set B includes problems that minimise figure and atmosphere effects.

Experiment 7: Statistical Summaries

Table C1.8

Experiment 7 Belief Bias Task ANCOVA Summary Table for Proportion of Positive ('Valid') Responses Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	.03	1	.03	.23	.64	.00
Error	12.91	91	.14			
Within-Subjects Effects						
OCD-Relevance	.05	1	.05	1.01	.32	.01
OCD-Relevance × OCD-Group	.07	1	.07	1.38	.24	.01
Error (OCD-Relevance)	4.48	92	.05			
Validity	4.62	1	4.62	64.29	<.001**	.41
Validity × OCD-Group	.15	1	.15	2.04	.16	.02
Error (Validity)	6.62	92	.07			
Believability	17.81	1	17.81	178.24	<.001**	.66
Believability × OCD-Group	.00	1	.00	.00	.97	.00
Error (Believability)	9.19	92	.10			
OCD-Relevance × Validity	.52	1	.52	11.95	.001**	.11
OCD-Relevance × Validity × OCD-Group	.00	1	.00	.00	.95	.00
Error (OCD-Relevance × Validity)	3.99	92	.04			
OCD-Relevance × Believability	1.03	1	1.03	29.06	<.001**	.24
OCD-Relevance × Believability × OCD-Group	.01	1	.01	.30	.58	.00
Error (OCD-Relevance × Believability)	3.25	92	.04			
Validity × Believability	.01	1	.01	.32	.57	.00
Validity × Believability × OCD-Group	.01	1	.01	.15	.70	.00
Error (Validity × Believability)	3.89	92	.04			
OCD-Relevance × Validity × Believability	.16	1	.16	4.35	.04*	.05
OCD-Relevance × Validity × Believability × OCD-Group	.07	1	.07	1.74	.19	.02
Error (OCD-Relevance × Validity × Believability)	3.46	92	.04			

Note. *p < .05, **p < .01.

Table C1.9

Experiment 7 Belief Bias Task ANCOVA Summary Table for Confidence Ratings Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Group	2.66	1	2.66	.86	.36	.01
Error	281.65	91	3.10			
Within-Subjects Effects						
OCD-Relevance	19.50	1	19.50	79.00	<.001**	.46
OCD-Relevance × OCD-Group	.04	1	.04	.17	.68	.00
Error (OCD-Relevance)	22.71	92	.25			
Validity	4.74	1	4.74	23.05	<.001**	.20
Validity × OCD-Group	.70	1	.70	3.42	.07	.04
Error (Validity)	18.93	92	.21			
Believability	5.19	1	5.19	15.69	<.001**	.15
Believability × OCD-Group	.21	1	.21	.62	.43	.01
Error (Believability)	30.43	92	.33			
OCD-Relevance × Validity	.01	1	.01	.03	.87	.00
OCD-Relevance × Validity × OCD-Group	.14	1	.14	.73	.39	.01
Error (OCD-Relevance × Validity)	17.63	92	.19			
OCD-Relevance × Believability	.55	1	.55	2.45	.12	.03
OCD-Relevance × Believability × OCD-Group	.01	1	.01	.03	.87	.00
Error (OCD-Relevance × Believability)	20.53	92	.22			
Validity × Believability	.94	1	.94	5.76	.02*	.06
Validity × Believability × OCD-Group	.27	1	.27	1.68	.20	.02
Error (Validity × Believability)	14.96	92	.16			
OCD-Relevance × Validity × Believability	.01	1	.01	.07	.79	.00
OCD-Relevance × Validity × Believability × OCD-Group	.14	1	.14	.82	.37	.01
Error (OCD-Relevance × Validity × Believability)	15.40	92	.17			

Note. *$p < .05$, **$p < .01$.

Table C1.10

Experiment 7 Belief Bias Task Mean Proportion of Positive ('Valid') Responses by OCD-Quartile

	OCD-neutral		OCD-relevant	
Item Type	Lowest OCD quartile	Highest OCD quartile	Lowest OCD quartile	Highest OCD quartile
Valid/Believable	0.88 (0.22)	0.77 (0.25)	0.76 (0.27)	0.70 (0.24)
Valid/Unbelievable	0.57 (0.26)	0.48 (0.27)	0.53 (0.33)	0.50 (0.27)
Invalid/Believable	0.64 (0.27)	0.68 (0.25)	0.61 (0.29)	0.63 (0.27)
Invalid/Unbelievable	0.34 (0.21)	0.26 (0.24)	0.42 (0.27)	0.48 (0.21)

Table C1.11

Experiment 7 Belief Bias Task ANCOVA Summary Table for Proportion of Positive ('Valid') Responses by OCD-Quartile, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
OCD-Quartile	.00	1	.00	.00	.99	.00
Error	7.04	51	.14			
Within-Subjects Effects						
OCD-Relevance	.00	1	.00	.03	.86	.00
OCD-Relevance × OCD-Quartile	.08	1	.08	1.85	.18	.03
Error (OCD-Relevance)	2.32	52	.04			
Validity	2.13	1	2.13	30.74	<.001**	.37
Validity × OCD-Quartile	.19	1	.19	2.78	.10	
Error (Validity)	3.60	52	.07			
Believability	7.24	1	7.24	70.61	<.001**	.58
Believability × OCD-Quartile	.00	1	.00	.02	.89	.00
Error (Believability)	5.33	52	.10			
OCD-Relevance × Validity	.33	1	.33	7.23	.01*	.12
OCD-Relevance × Validity × OCD-Quartile	.00	1	.00	.02	.90	.00
Error (OCD-Relevance × Validity)	2.38	52	.05			
OCD-Relevance × Believability	.51	1	.51	13.25	.001**	.20
OCD-Relevance × Believability × OCD-Quartile	.05	1	.05	1.19	.28	.02
Error (OCD-Relevance × Believability)	2.00	52	.04			
Validity × Believability	.00	1	.00	.03	.86	.00
Validity × Believability × OCD-Quartile	.04	1	.04	.72	.40	.01
Error (Validity × Believability)	2.62	52	.05			
OCD-Relevance × Validity × Believability	.08	1	.08	1.73	.19	.03
OCD-Relevance × Validity × Believability × OCD-Quartile	.03	1	.03	.72	.40	.01
Error (OCD-Relevance × Validity × Believability)	2.38	52	.05			

Note. *p < .05, **p < .01.

Appendix D

Chapter 5: Deductive and Inductive Reasoning in a Clinically Diagnosed Sample

Experiment 8

Experiment 8: Item Set

Table D1. 1

Experiment 8 Premise Monotonicity Task Neutral Stimuli

Validity	Premise 1	Premise 2	Premise 3	Premise 4	Premise 5	Conclusion
Valid	mammals	-	-	-	-	goats
	mammals	-	-	-	-	pigs
	mammals	-	-	-	-	monkeys
	mammals	-	-	-	-	zebras
	fruits	-	-	-	-	oranges
	fruits	-	-	-	-	bananas
	fruits	-	-	-	-	pineapples
	fruits	-	-	-	-	apricots
	flowers	-	-	-	-	roses
	flowers	-	-	-	-	daisies
	mammals	monkeys	lions	rabbits	mice	wolves
	mammals	rats	tigers	goats	cats	gorillas
	mammals	sheep	cows	pandas	goats	lions
	mammals	zebras	bears	foxes	monkeys	rabbits
	fruits	apricots	oranges	grapes	apples	strawberries
	fruits	cherries	bananas	raspberries	oranges	plums
	fruits	pears	peaches	mangoes	cherries	grapes
	fruits	grapes	blueberries	mandarins	peaches	limes
	flowers	violets	lilies	irises	lavenders	sunflowers
	flowers	roses	sunflowers	tulips	dandelions	poppies
Invalid	wolves	-	-	-	-	mammals
	deer	-	-	-	-	mammals
	lions	-	-	-	-	foxes
	dogs	-	-	-	-	horses
	cherries	-	-	-	-	fruits
	pears	-	-	-	-	fruits
	plums	-	-	-	-	peaches
	strawberries	-	-	-	-	grapes
	daffodils	-	-	-	-	flowers
	tulips	-	-	-	-	lavenders
	pandas	wolves	cows	horses	rats	mammals
	dogs	deer	gorillas	bears	cows	mammals
	cats	bears	foxes	zebras	mice	horses
	gorillas	foxes	tigers	pigs	zebras	cats
	pineapples	apples	bananas	watermelons	raspberries	fruits
	watermelons	mangoes	lemons	mandarins	nectarines	fruits
	rockmelons	strawberries	nectarines	cherries	mandarins	lemons
	pears	rockmelons	limes	apples	plums	watermelons
	frangipani	poppies	irises	tulips	orchids	flowers
	carnations	lavenders	violets	daisies	lilacs	dandelions

Note. The order of premises and item presentation were randomised for each participant.

Table D1.2

Experiment 8 Premise Monotonicity Task OCD-Relevant Stimuli

	Premise 1	Premise 2	Premise 3	Premise 4	Premise 5	Conclusion
Valid items						
All powered on/	household appliances	-	-	-	-	fans
All switched on	household appliances	-	-	-	-	vacuum cleaners
	household appliances	-	-	-	-	irons
	household appliances	-	-	-	-	washing machines
	household appliances	-	-	-	-	dishwashers
	household appliances	coffee machines	blenders	dishwashers	fans	ovens
	household appliances	sandwich presses	ovens	hair dryers	microwaves	heaters
	household appliances	vacuum cleaners	fans	clothes dryers	irons	electric kettles
	household appliances	air-conditioners	toasters	irons	blenders	clothes dryers
	household appliances	sandwich presses	microwaves	clothes dryers	electric kettles	deep fryers
All filthy/	kitchen products	-	-	-	-	knives
All dirty	kitchen products	-	-	-	-	measuring cups
	kitchen products	-	-	-	-	bowls
	kitchen products	-	-	-	-	teaspoons
	kitchen products	-	-	-	-	pots
	kitchen products	frying pans	forks	ladles	pots	teaspoons
	kitchen products	ladles	teaspoons	bowls measuring	saucepans	whisks
	kitchen products	wooden spoons	tablespoons	cups	frying pans	mugs
	kitchen products	chopping boards	plates	knives	teacups	wooden spoons
	kitchen products	measuring cups	knives	wooden spoons	mugs	tongs

Table D1.2

Continued

	Premise 1	Premise 2	Premise 3	Premise 4	Premise 5	Conclusion
Invalid items						
All powered on/	heaters	-	-	-	-	household appliances
All switched on	hair dryers	-	-	-	-	household appliances
	air-conditioners	-	-	-	-	microwaves
	deep fryers	-	-	-	-	toasters
	clothes dryers	-	-	-	-	sandwich presses
	coffee machines	electric kettles	fans	deep fryers	heaters	household appliances
	deep fryers	vacuum cleaners	heaters	electric kettles	toasters	household appliances
	vacuum cleaners	sandwich presses	hair dryers	microwaves	heaters	coffee machines
	ovens	clothes dryers	toasters	dishwashers	electric kettles	air-conditioners
	air-conditioners	coffee machines	deep fryers	irons	dishwashers	hair dryers
All filthy/	frying pans	-	-	-	-	kitchen products
All dirty	mugs	-	-	-	-	kitchen products
	tongs	-	-	-	-	chopping boards
	plates	-	-	-	-	forks
	tea cups	-	-	-	-	tablespoons
	frying pans	chopping boards	whisks	mugs	tablespoons	kitchen products
	saucepans	ladles	teacups	pots	tongs	kitchen products
	chopping boards	forks	pots	plates	teacups	bowls
	measuring cups	wooden spoons	plates	bowls	saucepans	ladles
	mugs	knives	teacups	teaspoons	whisks	saucepans

Note. The order of premises and item presentation were randomised for each participant.

Table D1.3

Depression, Anxiety and Stress Scales – 21 Item Version

1. I found it hard to wind down.
2. I was aware of dryness of my mouth.
3. I couldn't seem to experience any positive feeling at all.
4. I experienced breathing difficulty (e.g., excessively rapid breathing, breathlessness in the absence of physical exertion).
5. I found it difficult to work up the initiative to do things.
6. I trended to over-react to situations.
7. I experienced trembling (e.g., in the hands).
8. I felt that I was using a lot of nervous energy.
9. I was worried about situations in which I might panic and make a fool of myself.
10. I felt that I had nothing to look forward to.
11. I felt myself getting agitated.
12. I found it difficult to relax.
13. I felt down-hearted and blue.
14. I was intolerant of anything that kept me from getting on with what I was doing.
15. I felt I was close to panic.
16. I was unable to become enthusiastic about anything.
17. I felt I wasn't worth much as a person.
18. I felt that I was rather touchy.
19. I was aware of the action of my heart in the absence of physical exertion (e.g., sense of heart rate increase, heart missing a beat).
20. I felt scared without any good reason.
21. I felt that live was meaningless.

Note. Depression scale: items 3, 5, 10, 13, 16, 17, 21; Anxiety scale: items 2, 4, 7, 9, 15, 19, 20; Stress scale: items 1, 6, 8, 11, 12, 14, 18. Taken from www.psy.unsw.edu.au/dass/

Experiment 8: Statistical Summaries

Table D1.4

OCI-R Comparisons for All Analogue Samples Relative to the Clinical Sample

Study	Clinical cut-offs					Quartile groups				
	M	*N*	*t*	*df*	*p*	*M*	*N*	*t*	*df*	*p*
1	30.57 (8.21)	63	3.24	86	.002**	33.38 (7.67)	47	1.66	34.32	.11
2	32.24 (8.27)	46	2.29	69	.03	36.03 (6.92)	31	0.66	36.10	.51
3/6	29.95 (7.19)	44	2.95	33.66	.006**	34.00 (6.30)	27	1.41	35.30	.17
4/7	32.02 (8.59)	86	2.69	109	.008**	38.67 (10.63)	43	-0.31	34.01	.76
5	32.00 (10.63)	70	2.27	93	.03	-	-			

Note. ** $p < .01$. A significance value of .01 was assumed to control for the family wise error-rate.

Table D1.5

Experiment 8 Premise Monotonicity Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive ('Strong' / 'Valid') Responses, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.00	1	.00	.00	.95	.00
Error	12.01	47	.26			
Within-Subjects Effects						
OCD-Relevance	.40	1	.40	9.18	.004**	.16
OCD-Relevance × Clinical Group	.00	1	.00	.03	.87	.00
Error (OCD-Relevance)	2.07	48	.04			
Reasoning Condition	.73	1	.73	6.91	.011*	.13
Reasoning Condition × Clinical Group	.31	1	.31	2.95	.09	.06
Error (Reasoning Condition)	5.08	48	.11			
Validity	75.52	1	75.52	288.91	<.001**	.86
Validity × Clinical Group	.24	1	.24	.91	.35	.02
Error (Validity)	12.55	48	.26			
Length	2.18	1	2.18	18.42	<.001	.28
Length × Clinical Group	.02	1	.02	.19	.67	.00
Error (Length)	5.69	48	.12			
OCD-Relevance × Reasoning Condition	.01	1	.01	.34	.56	.01
OCD-Relevance × Reasoning Condition × Clinical Group	.00	1	.00	.10	.76	.00
Error (OCD-Relevance × Reasoning Condition)	1.20	48	.02			
Reasoning Condition × Validity	.22	1	.22	2.59	.11	.05
Reasoning Condition × Validity × Clinical Group	.04	1	.04	.49	.49	.01
Error (Reasoning Condition × Validity)	4.16	48	.09			
Reasoning Condition × Length	.99	1	.99	21.04	<.001**	.30
Reasoning Condition × Length × Clinical Group	.08	1	.08	1.61	.21	.03
Error (Reasoning Condition × Length)	2.27	48	.05			
OCD-Relevance × Validity	.03	1	.03	1.07	.31	.02
OCD-Relevance × Validity × Clinical Group	.02	1	.02	.73	.40	
Error (Content × Validity)	1.18	48	.02			
OCD-Relevance × Length	.21	1	.21	9.57	.003**	.17
OCD-Relevance × Length × Clinical Group	.03	1	.03	1.2	.28	.02
Error (OCD-Relevance × Length)	1.06	48	.02			
Validity × Length	5.61	1	5.61	80.50	<.001**	.63
Validity × Length × Clinical Group	.00	1	.00	.01	.94	.00
Error (Validity × Length)	3.35	48	.07			

Note. **p* < .05, ***p* < .01.

Table D1.6

Experiment 8 Premise Monotonicity Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive ('Strong' / 'Valid') Responses to Valid Arguments, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.23	1	.23	4.36	.04*	.08
Error	2.45	47	.05			
Within-Subjects Effects						
OCD-Relevance	.11	1	.11	2.90	.10	.06
OCD-Relevance × Clinical Group	.00	1	.00	.13	.72	.00
Error (OCD-Relevance)	1.80	48	.04			
Reasoning Condition	.07	1	.07	1.38	.25	.03
Reasoning Condition × Clinical Group	.06	1	.06	1.18	.28	.02
Error (Reasoning Condition)	2.54	48	.05			
Length	.40	1	.40	9.67	.003**	.17
Length × Clinical Group	.01	1	.01	.20	.66	.00
Error (Length)	1.97	48	.04			
OCD-Relevance × Reasoning Condition	.02	1	.02	.86	.36	.02
OCD-Relevance × Reasoning Condition × Clinical Group	.00	1	.00	.10	.76	.00
Error (OCD-Relevance × Reasoning Condition)	1.25	48	.03			
OCD-Relevance × Length	.30	1	.30	10.01	.003**	.17
OCD-Relevance × Length × Clinical Group	.06	1	.06	2.07	.16	.04
Error (Content × Length)	1.45	48	.03			
Reasoning Condition × Length	.08	1	.08	3.04	.09	.06
Reasoning Condition × Length × Clinical Group	.00	1	.00	.18	.68	.00
Error (Reasoning Condition × Length)	1.33	48	.03			
OCD-Relevance × Reasoning Condition × Length	.00	1	.00	.14	.71	.00
OCD-Relevance × Reasoning Condition × Length × Clinical Group	.05	1	.05	3.02	.09	.06
Error (OCD-Relevance × Reasoning Condition × Length)	.84	48	.02			

Note. **p* < .05, ***p* < .01.

Table D1.7

Experiment 8 Premise Monotonicity Induction-Deduction Task ANCOVA Summary Table for Proportion of Positive ('Strong' / 'Valid') Responses to Invalid Arguments, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.18	1	.18	.41	.53	.01
Error	21.15	47	.45			
Within-Subjects Effects						
OCD-Relevance	.31	1	.31	10.37	.002**	.18
OCD-Relevance × Clinical Group	.01	1	.01	.48	.49	.01
Error (OCD-Relevance)	1.45	48	.03			
Reasoning Condition	.88	1	.88	6.33	.02*	.12
Reasoning Condition × Clinical Group	.29	1	.29	2.09	.16	
Error (Reasoning Condition)	6.70	48	.14			
Length	7.40	1	7.40	50.25	<.001**	.51
Length × Clinical Group	.01	1	.01	.10	.76	.00
Error (Length)	7.07	48	.15			
OCD-Relevance × Reasoning Condition	.08	1	.08	3.53	.07	.07
OCD-Relevance × Reasoning Condition × Clinical Group	.01	1	.01	.65	.43	.01
Error (OCD-Relevance × Reasoning Condition)	1.07	48	.02			
OCD-Relevance × Length	.01	1	.01	.58	.45	.01
OCD-Relevance × Length × Clinical Group	.00	1	.00	.02	.88	.00
Error (Content × Length)	.83	48	.02			
Reasoning Condition × Length	1.25	1	1.25	17.19	<.001**	.26
Reasoning Condition × Length × Clinical Group	.10	1	.10	1.40	.24	.03
Error (Reasoning Condition × Length)	3.50	48	.07			
OCD-Relevance × Reasoning Condition × Length	.00	1	.00	.19	.67	.00
OCD-Relevance × Reasoning Condition × Length × Clinical Group	.00	1	.00	.19	.67	.00
Error (OCD-Relevance × Reasoning Condition × Length)	.93	48	.02			

Note. **p* < .05, ***p* < .01.

Table D1.8

Experiment 8 Premise Monotonicity Induction-Deduction Task ANCOVA Summary Table for Confidence Ratings for Valid Items, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	3.57	1	3.57	1.98	.17	.08
Error	84.93	47	1.81			
Within-Subjects Effects						
OCD-Relevance	.30	1	.30	1.85	.18	.04
OCD-Relevance × Clinical Group	.14	1	.14	.84	.37	.02
Error (OCD-Relevance)	7.85	48	.16			
Reasoning Condition	.01	1	.01	.02	.89	.00
Reasoning Condition × Clinical Group	.06	1	.06	.16	.69	.00
Error (Reasoning Condition)	18.97	48	.40			
Length	1.66	1	1.66	14.32	<.001**	.23
Length × Clinical Group	.20	1	.20	1.74	.19	.04
Error (Length)	5.58	48	.12			
OCD-Relevance × Reasoning Condition	.26	1	.26	2.31	.14	.05
OCD-Relevance × Reasoning Condition × Clinical Group	.02	1	.02	.20	.66	.00
Error (OCD-Relevance × Reasoning Condition)	5.40	48	.11			
OCD-Relevance × Length	.45	1	.45	6.44	.02*	.12
OCD-Relevance × Length × Clinical Group	.01	1	.01	.17	.68	.00
Error (Content × Length)	3.34	48	.07			
Reasoning Condition × Length	.02	1	.02	.23	.64	.00
Reasoning Condition × Length × Clinical Group	.00	1	.00	.03	.88	.00
Error (Reasoning Condition × Length)	4.78	48	.10			
OCD-Relevance × Reasoning Condition × Length	.01	1	.01	.10	.75	.00
OCD-Relevance × Reasoning Condition × Length × Clinical Group	.08	1	.08	.69	.41	.01
Error (OCD-Relevance × Reasoning Condition × Length)	5.83	48	.12			

Note. *p < .05, **p < .01.

Table D1.9

Experiment 8 Premise Monotonicity Induction-Deduction Task ANCOVA Summary Table for Confidence Ratings for Invalid Items, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	4.22	1	4.22	1.19	.28	.02
Error	166.22	47	3.54			
Within-Subjects Effects						
OCD-Relevance	.10	1	.10	1.20	.28	.02
OCD-Relevance × Clinical Group	.04	1	.04	.47	.50	.01
Error (OCD-Relevance)	4.10	48	.09			
Reasoning Condition	1.80	1	1.80	2.68	.11	.05
Reasoning Condition × Clinical Group	.14	1	.14	.22	.64	.00
Error (Reasoning Condition)	32.12	48	.67			
Length	7.62	1	7.62	19.76	<.001**	.29
Length × Clinical Group	.77	1	.77	2.01	.16	.04
Error (Length)	18.51	48	.39			
OCD-Relevance × Reasoning Condition	.04	1	.04	.38	.54	.01
OCD-Relevance × Reasoning Condition × Clinical Group	.04	1	.04	.38	.54	.01
Error (OCD-Relevance × Reasoning Condition)	5.12	48	.11			
OCD-Relevance × Length	.01	1	.01	.21	.65	.00
OCD-Relevance × Length × Clinical Group	.07	1	.07	1.42	.24	.03
Error (Content × Length)	2.28	48	.05			
Reasoning Condition × Length	.04	1	.04	.28	.60	.01
Reasoning Condition × Length × Clinical Group	.58	1	.58	4.01	.05	.08
Error (Reasoning Condition × Length)	6.92	48	.14			
OCD-Relevance × Reasoning Condition × Length	.21	1	.21	2.00	.16	.04
OCD-Relevance × Reasoning Condition × Length × Clinical Group	.00	1	.00	.00	.95	.00
Error (OCD-Relevance × Reasoning Condition × Length)	5.07	48	.11			

Note. **p* < .05, ***p* < .01.

Table D1.10

Experiment 8 Premise Monotonicity Task Mean Proportion of Positive ('Strong'/ 'Valid') Responses to Invalid Items by Conclusion Type

		General Conclusions		Specific Conclusions	
Reasoning Condition	Argument Length	Low OCD group	High OCD group	Low OCD group	High OCD group
OCD-neutral items					
Deduction	One-Premise	0.17 (0.35)	0.22 (0.36)	0.09 (0.24)	0.11 (0.30)
	Five-Premise	0.37 (0.45)	0.37 (0.48)	0.29 (0.39)	0.24 (0.35)
Induction	One-Premise	0.07 (0.17)	0.21 (0.39)	0.08 (0.17)	0.12 (0.27)
	Five-Premise	0.48 (0.46)	0.65 (0.40)	0.36 (0.42)	0.40 (0.46)
OCD-relevant items					
Deduction	One-Premise	0.23 (0.42)	0.22 (0.39)	0.14 (0.31)	0.14 (0.30)
	Five-Premise	0.40 (0.48)	0.32 (0.45)	0.37 (0.46)	0.28 (0.39)
Induction	One-Premise	0.12 (0.30)	0.21 (0.38)	0.15 (0.29)	0.24 (0.38)
	Five-Premise	0.59 (0.46)	0.73 (0.42)	0.46 (0.45)	0.49 (0.48)

Table D1.11

Experiment 8 Premise Monotonicity Task ANCOVA Summary Table for Proportion of Positive ('Strong' / 'Valid') Responses for Invalid Items by Conclusion Type, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.44	1	.44	.49	.49	.01
Error	42.43	47	.90			
Within-Subjects Effects						
OCD-Relevance	.58	1	.58	9.38	.004**	.16
OCD-Relevance × Clinical Group	.04	1	.04	.61	.44	.01
Error (OCD-Relevance)	2.96	48	.06			
Reasoning Condition	1.55	1	1.55	5.05	.03*	.10
Reasoning Condition × Clinical Group	.63	1	.63	2.07	.16	.04
Error (Reasoning Condition)	14.70	48	.31			
Conclusion Type	1.55	1	1.55	10.66	.002**	.18
Conclusion Type × Clinical Group	.14	1	.14	.95	.34	.02
Error (Conclusion Type)	6.96	48	.15			
Length	14.40	1	14.40	48.34	<.001**	.50
Length × Clinical Group	.06	1	.06	.19	.67	.00
Error (Length)	14.30	48	.30			
OCD-Relevance × Reasoning Condition	.13	1	.13	2.81	.10	.06
OCD-Relevance × Reasoning Condition × Clinical Group	.02	1	.02	.40	.53	.01
Error (OCD-Relevance × Reasoning Condition)	2.21	48	.05			
OCD-Relevance × Conclusion Type	.08	1	.08	1.57	.22	.03
OCD-Relevance × Conclusion Type × Clinical Group	.03	1	.03	.60	.44	.01
Error (OCD-Relevance × Conclusion Type)	2.34	48	.05			

Table D1.11
Continued

Effect	SS	df	Mean Square	F	Sig	η_p^2
OCD-Relevance × Length	.01	1	.01	.16	.69	.00
OCD-Relevance × Length × Clinical Group	.01	1	.01	.13	.72	.00
Error (Content × Length)	2.59	48	.05			
Reasoning Condition × Conclusion Type	.01	1	.01	.08	.78	.00
Reasoning Condition × Conclusion Type × Clinical Group	.05	1	.05	.56	.46	.01
Error (Reasoning Condition × Conclusion Type)	4.55	48	.09			
Reasoning Condition × Length	2.14	1	2.14	15.33	<.001**	.24
Reasoning Condition × Length × Clinical Group	.07	1	.07	.48	.49	.01
Error (Reasoning Condition × Length)	6.69	48	.14			
Conclusion Type × Length	.33	1	.33	4.56	.04*	.09
Conclusion Type × Length × Clinical Group	.02	1	.02	.32	.57	.01
Error (Conclusion Type × Length)	3.51	48	.07			
OCD-Relevance × Reasoning Condition × Conclusion Type	.00	1	.00	.04	.83	.00
OCD-Relevance × Reasoning Condition × Conclusion Type × Clinical Group	.00	1	.00	.01	.93	.00
Error (OCD-Relevance × Reasoning Condition × Conclusion Type)	1.85	48	.04			
OCD-Relevance × Reasoning Condition × Length	.02	1	.02	.57	.46	.01
OCD-Relevance × Reasoning Condition × Length × Clinical Group	.00	1	.00	.01	.91	.00
Error (OCD-Relevance × Reasoning Condition × Length)	1.98	48	.04			
OCD-Relevance × Conclusion Type × Length	.00	1	.00	.00	1.00	.00
OCD-Relevance × Conclusion Type × Length × Clinical Group	.01	1	.01	.21	.65	.00
Error (OCD-Relevance × Conclusion Type × Length)	1.56	48	.03			
Reasoning Condition × Conclusion Type × Length	.50	1	.50	12.26	.001**	.20
Reasoning Condition × Conclusion Type × Length × Clinical Group	.01	1	.01	.22	.64	.00
Error (Reasoning Condition × Conclusion Type × Length)	1.96	48	.04			
OCD-Relevance × Reasoning Condition × Conclusion Type × Length	.05	1	.05	.83	.37	.02
OCD-Relevance × Reasoning Condition × Conclusion Type × Length × Clinical Group	.00	1	.00	.06	.80	.00
Error (OCD-Relevance × Reasoning Condition × Conclusion Type × Length)	2.61	48	.05			

Note. *$p < .05$, **$p < .01$.

Table D 1.12

Experiment 8 Forced-Choice Premise Diversity Task ANCOVA Summary Table for Proportion of Diverse Arguments Accepted, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.06	1	.06	.39	.53	.01
Error	7.16	47	.15			
Within-Subjects Effects						
OCD-Relevance	.06	1	.06	4.97	.03*	.09
OCD-Relevance × Clinical Group	.00	1	.00	.18	.68	.00
Error (OCD-Relevance)	.59	48	.01			

Note. *p < .05, **p < .01.

Table D1.13

Experiment 8 Forced-Choice Premise Diversity Task ANCOVA Summary Table for Perceived Usefulness Ratings, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.88	1	.88	.08	.77	.00
Error	497.16	47	10.58			
Within-Subjects Effects						
OCD-Relevance	5.25	1	5.25	9.48	.003**	.16
OCD-Relevance × Clinical Group	.87	1	.87	1.57	.22	.03
Error (OCD-Relevance)	26.58	48	.55			
Diversity	27.98	1	27.98	9.26	.004**	.16
Diversity × Clinical Group	16.51	1	16.51	5.47	.02*	.10
Error (Diversity)	144.94	48	3.02			
OCD-Relevance × Diversity	.36	1	.36	1.23	.27	.02
OCD-Relevance × Diversity × Clinical Group	.19	1	.19	.63	.43	.01
Error (Content × Diversity)	14.26	48	.30			

Note. *p < .05, **p < .01.

Table D1.14

Experiment 8 Similarity Rating Task ANCOVA Summary Table for Proportion of Matching Responses, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.00	1	.00	.61	.44	.01
Error	.18	47	.00			
Within-Subjects Effects						
OCD-Relevance	.05	1	.05	24.26	<.001**	.34
OCD-Relevance × Clinical Group	.00	1	.00	.02	.89	.00
Error (OCD-Relevance)	.11	48	.00			

Note. **p* < .05, ***p* < .01.

Table D1.15

Experiment 8 Similarity Rating Task ANCOVA Summary Table for Similarity Ratings, Controlling for General Ability

Effect	SS	df	Mean Square	F	Sig	η_p^2
Between-Subjects Effects						
Clinical Group	.02	1	.02	.00	.95	.00
Error	237.07	47	5.04			
Within-Subjects Effects						
OCD-Relevance	13.21	1	13.21	28.53	<.001**	.37
OCD-Relevance × Clinical Group	.01	1	.01	.01	.92	.00
Error (OCD-Relevance)	22.22	48	.46			
Diversity	705.13	1	705.13	217.34	<.001**	.82
Diversity × Clinical Group	.24	1	.24	.07	.79	.00
Error (Diversity)	155.73	48	3.24			
OCD-Relevance × Diversity	3.79	1	3.79	17.19	<.001**	.26
OCD-Relevance × Diversity × Clinical Group	.00	1	.00	.02	.90	.00
Error (Content × Diversity)	10.58	48	.22			

Note. **p* < .05, ***p* < .01.

Table D1.16

Experiment 8 Premise Monotonicity Task Estimated Marginal Mean Proportion of Positive ('Strong'/ 'Valid') Responses by Clinical Group (Standard Deviations in Parentheses)

		OCD		Non-Clinical Controls		Anxious Controls	
Validity	Length	Deduction	Induction	Deduction	Induction	Deduction	Induction
OCD-Neutral Items							
Valid	1P	0.94 (0.17)	0.95 (0.15)	0.94 (0.14)	0.95 (0.12)	0.94 (0.12)	0.91 (0.22)
	5P	0.72 (0.35)	0.86 (0.26)	0.86 (0.22)	0.87 (0.25)	0.87 (0.22)	0.83 (0.30)
Invalid	1P	0.17 (0.30)	0.14 (0.27)	0.13 (0.26)	0.07 (0.14)	0.26 (0.36)	0.24 (0.33)
	5P	0.29 (0.36)	0.53 (0.37)	0.32 (0.38)	0.53 (0.39)	0.40 (0.44)	0.61 (0.38)
OCD-Relevant Items							
Valid	1P	0.89 (0.16)	0.91 (0.14)	0.97 (0.08)	0.92 (0.16)	0.74 (0.37)	0.91 (0.25)
	5P	0.89 (0.16)	0.93 (0.14)	0.90 (0.23)	0.94 (0.18)	0.79 (0.30)	0.93 (0.24)
Invalid	1P	0.18 (0.29)	0.21 (0.31)	0.17 (0.33)	0.14 (0.24)	0.26 (0.36)	0.29 (0.35)
	5P	0.29 (0.38)	0.63 (0.38)	0.38 (0.43)	0.52 (0.40)	0.36 (0.39)	0.67 (0.38)

Table D1.17

Experiment 8 Premise Diversity Task Estimated Marginal Mean Proportion of Diverse Pairs Accepted and Perceived Usefulness Ratings (Standard Deviations in Parentheses)

	OCD-neutral			OCD-relevant		
Clinical group	Proportion of diverse pairs chosen	Diverse evidence usefulness rating	Non-diverse evidence usefulness rating	Proportion of diverse pairs chosen	Diverse evidence usefulness rating	Non-diverse evidence usefulness rating
OCD patient	0.70 (0.32)	5.67 (2.10)	5.41 (1.82)	0.66 (0.30)	6.10 (1.72)	5.95 (1.70)
Non-clinical control	0.76 (0.31)	6.28 (2.32)	4.81 (1.85)	0.70 (0.28)	6.32 (1.76)	5.15 (1.75)
Anxious control	0.67 (0.34)	6.10 (2.26)	5.41 (1.91)	0.69 (0.22)	6.47 (2.23)	5.50 (1.97)

Table D1.18

Experiment 8 Similarity Rating Task Estimated Marginal Mean Proportion of 'Matching' Responses (i.e., Nominally 'Non-Diverse' Items Endorsed as the More Similar Premises) and Similarity Ratings (Standard Deviations in Parentheses)

	OCD-neutral			OCD-relevant		
Clinical group	Proportion of matching responses	'Diverse pair' similarity rating	'Non-diverse pair' similarity rating	Proportion of matching responses	'Diverse pair' similarity rating	'Non-diverse pair' similarity rating
OCD patient	0.98 (0.03)	3.72 (1.45)	7.69 (0.92)	0.94 (0.06)	4.52 (1.56)	7.93 (1.01)
Non-clinical control	0.97 (0.05)	3.65 (2.05)	7.74 (1.66)	0.93 (0.07)	4.42 (1.82)	7.98 (1.28)
Anxious control	0.94 (0.13)	4.00 (2.36)	8.06 (1.14)	0.91 (0.13)	4.67 (1.80)	8.24 (0.93)