ABSTRACT

ORDERING AND ARRANGING BEHAVIORS IN OBSESSIVE-COMPULSIVE DISORDER: A REPLICATION AND EXTENSION

Sarah R. Lee, M.A.
Department of Psychology
Northern Illinois University, 2018
Kevin D. Wu, Director

In the existing Obsessive-Compulsive Disorder (OCD) research literature, some of its “symptom dimensions” have been studied more extensively than others. Symmetry and ordering concerns, for example, have been the subject of far less research than washing or checking symptoms. In one of the few extant studies of these concerns, researchers previously found that participants with elevated ordering concerns reported higher anxiety in response to a stressful task when assigned to a disorganized office space, as opposed to an organized one. The current study was intended as a replication of this prior study in a larger sample.

One hundred twenty-six undergraduate students from Northern Illinois University were recruited for this study and sorted into “high” and “low” ordering groups based on self-report scores of ordering concerns. Participants were randomized to work in either an organized or disorganized office space, and post-test anxiety ratings were gathered as the major dependent variable. Multiple regression was used to test the primary hypothesis that participants in the high ordering concerns/disorganized condition would be most anxious at post-test. Despite an increase in statistical power as compared to the original study, this hypothesis was not supported. Neither self-reported ordering concerns nor room condition (nor the interaction of these two variables) was found to contribute to post-test anxiety ratings. Study limitations are considered in detail.
Also considered are the implications of these findings for the study of OCD-ordering symptoms and for replication efforts of published research more broadly.
ORDERING AND ARRANGING BEHAVIORS IN OBSESSIVE-COMPULSIVE DISORDER:
A REPLICATION AND EXTENSION

BY

SARAH R. LEE
©2018 Sarah R. Lee

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF ARTS

DEPARTMENT OF PSYCHOLOGY

Thesis Director:
Kevin D. Wu
ACKNOWLEDGEMENTS

I am grateful to many individuals for their time and expertise in helping this project come to be. I would first like to thank my thesis advisor, Kevin Wu, who provided me with support through every step of this process. Without his encouragement along the daunting process of developing my own experimental project and his critical eye toward improving the strength of my writing, this project would not exist. I would also like to thank my thesis committee members, Karen White and David Valentiner, for their comments and support of this project. Finally, I would like to thank the research assistants who contributed hours to the data collection efforts for this study: Kelly, Katelyn, Tom, Sonya, Juan, Mike, and especially Sarah. I would never have gathered enough participants without your help. Thank you!
DEDICATION

To my parents for their ongoing support, and to Ryan, who was with me every step
TABLE OF CONTENTS

LIST OF TABLES........................................................................................................ vi
LIST OF APPENDICES.................................................................................................. vii

Chapter

1. INTRODUCTION........................................................................................................ 1
   A Cognitive-Behavioral Model of OCD................................................................. 3
   An Alternative Model....................................................................................... 10
   The Importance of Symptom Type: Relationship to Belief Domains............ 15
   The Present Study.......................................................................................... 21

2. METHODS............................................................................................................... 24
   Participants....................................................................................................... 24
   Measures........................................................................................................ 25
   Procedure........................................................................................................ 30

3. DATA ANALYSIS AND RESULTS........................................................................ 36
   Data Cleaning.................................................................................................. 36
   Preliminary Analyses..................................................................................... 38
   Primary Analyses.......................................................................................... 40

4. DISCUSSION......................................................................................................... 46
   Limitations........................................................................................................ 48
   Implications.................................................................................................... 53
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>56</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDICES</td>
<td>63</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Normality tests for predictor variables</td>
<td>37</td>
</tr>
<tr>
<td>2. Independent variable descriptive statistics</td>
<td>38</td>
</tr>
<tr>
<td>3. Average inter-item correlations of continuous measures</td>
<td>39</td>
</tr>
<tr>
<td>4. Group means on SUDS, full sample</td>
<td>41</td>
</tr>
<tr>
<td>5. Model summary, full sample, SUDS as outcome</td>
<td>42</td>
</tr>
<tr>
<td>6. Regression results, full sample, SUDS as outcome</td>
<td>42</td>
</tr>
<tr>
<td>7. Group means on SUDS, extreme sample</td>
<td>43</td>
</tr>
<tr>
<td>8. Model summary, extreme sample, SUDS as outcome</td>
<td>43</td>
</tr>
<tr>
<td>9. Regression results, extreme sample, SUDS as outcome</td>
<td>43</td>
</tr>
<tr>
<td>10. Group means, STICSA-S, full sample</td>
<td>44</td>
</tr>
<tr>
<td>11. Model summary, full sample, STICSA-S as outcome</td>
<td>44</td>
</tr>
<tr>
<td>12. Regression results, full sample, STICSA-S as outcome</td>
<td>44</td>
</tr>
</tbody>
</table>
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. DEMOGRAPHIC INFORMATION QUESTIONNAIRE</td>
<td>63</td>
</tr>
<tr>
<td>B. THE SCHEDULE FOR COMPULSIONS, OBSESSIONS AND PATHOLOGICAL IMPULSES—COMPULSIVE RITUALS SUBSCALE (SCOPI-CR)</td>
<td>65</td>
</tr>
<tr>
<td>C. THE SOCIAL INTERACTION ANXIETY SCALE AND THE SOCIAL PHOBIA SCALE</td>
<td>67</td>
</tr>
<tr>
<td>D. THE STATE-TRAIT INVENTORY FOR COGNITIVE AND SOMATIC ANXIETY—STATE VERSION</td>
<td>69</td>
</tr>
<tr>
<td>E. THE POSITIVE AND NEGATIVE AFFECT SCHEDULE—NEGATIVE AFFECT SCALE</td>
<td>71</td>
</tr>
<tr>
<td>F. ORDERING AND ARRANGING BEHAVIORS INTERVIEW</td>
<td>73</td>
</tr>
<tr>
<td>G. VALIDITY QUESTIONS</td>
<td>75</td>
</tr>
<tr>
<td>H. PROCEDURE</td>
<td>77</td>
</tr>
<tr>
<td>I. DISORGANIZED ROOM STIMULUS</td>
<td>79</td>
</tr>
<tr>
<td>J. ORGANIZED ROOM STIMULUS</td>
<td>81</td>
</tr>
<tr>
<td>K. DATA LOSS IN CURRENT STUDY</td>
<td>83</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

In the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association [APA], 2013), a diagnosis of Obsessive-Compulsive Disorder (OCD) is characterized by the presence of at least one of two features: obsessions and compulsions. Obsessions are thoughts, images, or urges which persist despite being experienced as intrusive and unwanted. Compulsions are repetitive behavioral or mental rituals that individuals feel compelled to perform in response to their obsessions or according to rigidly applied rules. If left untreated, OCD shows a chronic and sometimes deteriorating course (APA, 2013). Based on a nationally representative sample, the 12-month and lifetime prevalencies of OCD in the United States are estimated to be 1.2% and 2.3%, respectively, although 28.2% of participants indicated experiencing obsessions or compulsions at some point in their lives (Ruscio, Stein, Chiu, & Kessler, 2010). The World Health Organization has listed OCD among the leading causes of disability in adolescents and young adults (Kohn, Saxena, Levav, & Saraceno, 2004). Due to the large percentages of people who experience sub-syndromal symptoms, the disease burden of OCD is likely underestimated (Ruscio et al., 2010). Thus, OCD represents a source of significant distress and impairment to afflicted individuals around the world.

Research and clinical observations show that OCD is a heterogeneous disorder. It is possible for individuals to meet criteria for the disorder with (a) obsessions in the absence of
compulsions, (b) compulsions in the absence of obsessions, or (c) obsessions and compulsions together (APA, 2013). What is more, the content of obsessions can vary widely, including although not limited to religious themes, themes of violence, and sexual themes (Rachman, 1998). Common compulsions include arranging items, counting, washing objects or one’s self, and checking to be sure that lights are off or doors are locked (Rachman, 2002; Salkovskis, 1985). Symptoms among children are similar, with some of the most common symptoms including aggressive or sexual obsessions, superstitions, and order-related obsessions (McKay et al., 2006). Although clients may present with these symptoms in any combination, it is not uncommon practice among OCD researchers to limit participants to endorsing only one “primary” symptom domain. When such restrictions are removed, over 50% of participants rate more than one type of symptom as their most severe, suggesting that perhaps multiple symptoms are being experienced as equally severe (Gönner, Ecker, & Leonhart, 2009).

This practice of limiting participants to one primary symptom has resulted in research efforts focused on some symptom dimensions (especially washing and checking) more than others (such as ordering, neutralizing, hoarding, obsessing; McKay et al., 2004). In their review of evidence for symptom subtypes, McKay and colleagues (2004) noted that clients with primary washing or checking symptoms made up 75% of the research samples and that other kinds of symptoms were severely underrepresented. This differential emphasis in research has implications for clinical practice, where primary symptom domain seems to influence response to therapy. As an example, one study found that although only 20% of clients with obsessions without compulsions show reliable clinical change (which authors defined as a Yale-Brown Obsessive-Compulsive Scale [Y-BOCS; Goodman et al., 1989] score that decreased by 6
points from the pretreatment score and was less than 12 points overall), up to 33% of clients with checking behaviors show such change (McLean et al., 2001). In another example, when reviewing the recent literature on exposure and response prevention treatments for OCD, Williams, Mugno, Franklin, and Faber (2013) concluded that ordering and arranging symptoms seem to respond worse to exposure and response prevention therapy than do washing or checking symptoms. However, it is important to note the relative paucity of published research on treatment of ordering symptoms, thus rendering this conclusion as a tentative one. In short, symptoms which have been the focus of research appear to respond better in therapy (McKay et al., 2004; McLean et al., 2001).

The present study examines ordering and arranging concerns, one of the symptom domains that has received relatively less emphasis within the contemporary OCD literature. It is especially noteworthy that this symptom domain has been understudied given its relatively high prevalence in clinical samples (26.6% of patients in Denys, de Geus, van Megen, & Westenberg, 2004). This research examines ordering and arranging concerns in light of the cognitive-behavioral model of OCD.

A Cognitive-Behavioral Model of OCD

Cognitive processes have held a place of importance in understanding OCD for some time and are important to the present study. Salkovskis (1985) articulated the first comprehensive cognitive-behavioral model of OCD, synthesizing the work of many prominent behavioral OCD researchers of the time with a cognitive theory similar to those that were being
formulated for other disorders. This model attempted to explain the finding that much of the general population experiences obsession-like thoughts (Rachman & de Silva, 1978), but only a minority of people develop OCD. In Salkovskis’s (1985) model, the way people interpret their intrusive thoughts is more important than the content of those thoughts. Intrusive thoughts are experienced by all people as ego-dystonic, or in conflict with their self-image, but if people attribute no importance to them, a self-perpetuating cycle of intrusive thoughts and attempts to neutralize those thoughts will not occur (Salkovskis, 1985). Interpreting intrusive thoughts as threatening, Salkovskis argued, will lead to the individual experiencing anxiety; interpreting intrusive thoughts with feelings of loss will lead to depression. For a thought to become obsessional, the thinker must interpret it with feelings of personal blame or responsibility (Salkovskis, 1985).

The importance of responsibility to OCD seems to have been confirmed by findings from Taylor and colleagues (2010), in which responsibility was found to predict all investigated domains of obsessive-compulsive symptoms. Feelings of responsibility lead the individual to neutralize an intrusive thought using overt (compulsive behaviors) or covert (mental neutralizing) strategies (Salkovskis, 1985). These neutralizing strategies are meant to prevent the feared consequence of the thought from occurring because the individual will feel directly or indirectly responsible for those consequences. The neutralizing rituals decrease discomfort in the short term, and often the feared consequence does not occur; both of these serve to reinforce the neutralizing behavior (Salkovskis, 1985).
According to this model, discomfort and subsequent urges to neutralize will be increased as either the perceived likelihood or the perceived awfulness of the consequence increases (Salkovskis, 1999). A client may recognize the unlikelihood of the feared event yet have powerful urges to neutralize anyway, due to the extreme awfulness of the feared event (Salkovskis, 1999). Because the cognitive-behavioral model states that level of neutralization is contingent upon the way intrusive thoughts are interpreted, not the content of intrusive thoughts themselves, recommendations for therapy in addition to behavioral exposure and response prevention could involve challenging clients’ interpretations and coming up with more benign alternatives, instead of challenging the intrusions directly (Salkovskis, 1985, 1999).

Rachman (2002) described what this cognitive-behavioral model would look like in an individual with checking compulsions. Like Salkovskis (1985), he emphasized the importance of responsibility in the development and maintenance of symptoms; the individual feels that he or she has some kind of elevated responsibility for preventing harm (Rachman, 2002). To prevent harm, the individual checks for safety (e.g., that a door is locked, preventing burglary; that they have not hit anyone with their car, preventing injury), wanting to feel certain that the probability of harm has been removed entirely. However, because it is impossible to predict future events with perfect certainty, and because these fears of harm are often general or vague, the checking persists without a natural end (Rachman, 2002). In addition, the more times the individual checks, the less confident one becomes in one’s memory of having checked (e.g., “Did I lock the door?” [Radomsky & Alcolado, 2010]), causing the person to check again in case he or she did not before (Rachman, 2002).
Behavioral Experiments

Subsequent research has refined and enhanced Salkovskis’s (1985) model. In a now-classic experiment, Lopatka and Rachman (1995) attempted to manipulate the responsibility felt by subjects with a diagnosis of OCD during a behavioral approach test (BAT) conducted in the subjects’ own homes. Participants were individuals with clinical diagnoses of OCD: 30 with primary checking behavior and 10 with primary cleaning behavior. The BAT brought participants face to face with stimuli that would induce their primary obsessions, in this experiment, urges to check or urges to clean. Performing the BAT in the home made the stimuli personally salient, and the researchers manipulated feelings of responsibility for the situation through use of contracts (Lopatka & Rachman, 1995). In the low-responsibility condition, the contracts signed by the researcher and the participant stated that what happened as a result of the experimental procedure was the sole responsibility of the researcher; in the high-responsibility condition, the contracts stated that whatever happened was the sole responsibility of the participant; in the control condition, the contract did not specify who was held responsible for the situation. Consistent with the cognitive-behavioral model of OCD, feelings of participant responsibility ($t = 6.52, p < .01$), urges to check ($t = 4.35, p < .01$), and feelings of discomfort ($t = 4.74, p < .01$) were significantly lowered by the researcher taking responsibility for consequences, compared to the control condition in which responsibility was unclear (Lopatka & Rachman, 1995). The feelings of responsibility and discomfort, as well as urges to check, were not significantly higher in the high-responsibility condition compared to the control condition; since responsibility was not specified to either the researcher or participants in the control condition, participants may have dealt with this ambiguity by taking the responsibility upon
themselves (Lopatka & Rachman, 1995).

Other researchers have attempted to replicate Lopatka and Rachman’s (1995) findings. Moulding, Kyrios, and Doron (2007) presented four vignettes that varied in level of responsibility (high or low) and level of threat (high or low) to 219 university students and asked them to imagine they were actually experiencing these situations. The researchers then followed up with questions about their discomfort levels and urges to do something. Moulding and colleagues (2007) found that urges to take action \( (F(3, 218) = 92.19, p < .007) \), as well as feelings of discomfort \( (F(3, 218) = 265.32, p < .007) \), were significantly higher in the high-responsibility conditions as compared to the low-responsibility conditions. Urges to take action \( (F(3, 218) = 70.56, p < .007) \) and feelings of discomfort \( (F(3, 218) = 666.62, p < .007) \) were also significantly higher in the high-threat conditions as compared to the low-threat conditions (Moulding et al., 2007). These findings suggest that both perceived responsibility and perceived threat significantly increase feelings of discomfort and urges to do something in response to discomfort in a sample of college students.

Another replication performed by Arntz, Voncken, and Goosen (2007) provided a more practical re-application of Lopatka and Rachman’s (1995) ideas. This replication was performed using people diagnosed with OCD, a group of anxious control patients, and a group of non-patients using an actual, instead of imagined, scenario. Unlike Lopatka and Rachman (1995), who performed their behavioral experiment in participant homes, Arntz and colleagues (2007) presented all participants with a novel pill-sorting task in the laboratory. They created high-and low-responsibility conditions by either emphasizing the importance and potential impact of the pill sort or not mentioning its importance or impact at all. As they hypothesized, participants exposed to the high-responsibility condition reported significantly higher danger ratings.
(F(1, 86)= 62.31, \( p < .001 \)), with participants with OCD reporting danger levels which did not differ significantly from non-patients or anxious controls. However, the high responsibility seemed to affect their behaviors more, with the OCD group performing significantly more checking behaviors than other groups (F(2, 83) = 6.07, \( p = .003 \)) and taking significantly longer to complete the pill-sorting task (F(2, 83) = 4.16, \( p = .019 \)). Thus, Arntz and colleagues (2007) were able to successfully replicate Lopatka and Rachman’s (1995) findings.

Based upon Lopatka and Rachman’s (1995) study, as well as a subsequent and unsuccessful manipulation of their own, Radomsky and Rachman (2004a) noted an additional factor for consideration in the cognitive behavioral model: personal relevance of stimuli. Radomsky and Rachman (2004a) had attempted to conduct an experiment examining memory bias in OCD. They presented groups of college students with higher than average or lower than average ordering and arranging concerns with organized or disorganized workspaces, hypothesizing that participants with higher ordering and arranging concerns would have better memory for disorganized objects than participants with lower ordering and arranging concerns. However, this investigation found no significant effects, either of subjects’ ordering concerns or room condition. As a possible explanation for their findings, Radomsky and Rachman (2004a) suggested that stimuli must be personally relevant to cause any effects, and their stimuli, placed in an office of the psychology building with no significance to undergraduate students, did not meet this criterion. This issue of personal relevance of stimuli could also be why Lopatka and Rachman (1995), who conducted their procedures in participant homes, experienced such success. However, another possible issue in Radomsky and Rachman’s (2004a) study may have been the difficulty of the memory task. In an entire room of items, participants were asked to
recall four items specifically, and the study may have suffered floor effects due to this difficulty.

**Additional Cognitive Biases**

Rachman (1997) expanded on the cognitive-behavioral model by describing cognitive biases besides responsibility likely to play a role in the experience of obsessions. These other cognitive biases include a belief that things are more likely to go wrong when one is responsible (thus increasing the likelihood of feared consequences) and a belief called *thought-action fusion*. As Rachman (1997) described, thought-action fusion can take two forms: (a) a belief that having a thought about a feared event can increase its likelihood and (b) a belief that thinking of a repulsive action is morally equivalent to performing that action. Thought-action fusion may be a product of and also serve to increase responsibility in affected individuals (Rachman, 1997).

Rachman (1998) later stated that the obsessions one has typically reflect values that are most important to the individual, echoed by Salkovskis’s (1999) assertion that highly religious people are more likely to have and be bothered by blasphemous thoughts, and peaceful people are more likely to have and be bothered by violent thoughts. In addition, Rachman (1998) also asserted that the number of intrusions increases during times of stress, which can result in a vicious cycle in which intrusions exacerbate stress, and stress increases the number of intrusions.

**OCCWG Belief Domains**

The Obsessive-Compulsive Cognitions Working Group (OCCWG, 1997), a panel of
OCD research experts, expanded the cognitive-behavioral model further by suggesting five belief domains in addition to responsibility which researchers deemed important in OCD: overestimation of the importance of one’s thoughts, excessive concern with controlling one’s thoughts, overestimation of potential threats, intolerance for feelings of uncertainty, and a tendency toward perfectionism. In a recent review of the 30 years of research which have passed since the initial articulation of the cognitive-behavioral model, Salkovskis and Millar (2016) assert that, whereas each of these six belief domains is relevant to OCD, only some are specific to the disorder. They view overestimation of threat, perfectionism, and intolerance of uncertainty as dysfunctional beliefs which leave one vulnerable to developing a variety of disorders (Salkovskis & Millar, 2016). As an example, intolerance of uncertainty is a construct which was originally devised as part of a cognitive model explaining Generalized Anxiety Disorder (Dugas, Gagnon, Ladouceur, & Freeston, 1998) rather than OCD. Meanwhile, only inflated responsibility, overestimation of the importance of thoughts, and excessive concern with controlling one’s thoughts are believed to be specific to OCD (Salkovskis & Millar, 2016). Nonetheless, the perspective of the cognitive-behavioral model has widened considerably from the initial focus on just responsibility beliefs (Salkovskis, 1985).

An Alternative Model

Rachman (2002) described that responsibility beliefs in the prevailing cognitive-behavioral model involved a responsibility for preventing harm. With this emphasis on the importance of avoiding harm, the cognitive-behavioral model for OCD was similar to those of
other anxiety disorders (Summerfeldt, 2004). However, though much research focused on the harm-avoidant function of compulsions, Summerfeldt (2004) created an alternate framework describing motivations of obsessive-compulsive symptoms. According to Summerfeldt’s (2004) assertions, responsibility is one motivator for OCD, but feelings of incompleteness are another; these are believed to be two continuous, orthogonal constructs. Researchers created a measure, the Obsessive-Compulsive Traits Core Dimensions Questionnaire (OC-TCDQ; Summerfeldt, Kloosterman, Parker, Antony, & Swinson, 2001), to capture both of these traits. The scale contains 20 items to assess harm avoidance and incompleteness (Summerfeldt et al., 2001). To support Summerfeldt’s (2004) assertion that these should be two separable constructs, researchers attempted to subject OC-TCDQ to confirmatory factor analysis (Pietrefesa & Coles, 2008). Pietrefesa and Coles’s analysis examined an obliquely rotated two-factor solution in comparison to a forced one-factor solution using maximum likelihood estimation. The nested models were compared, confirming that the two-factor solution was superior to the one-factor solution ($\chi^2_{\text{diff}} = 40.39, p < .001$). Items were specified to load on a single factor, so cross-loadings were not reported; item loadings ranged from .60 to .79 on factor 1 ($p$’s < .001) and from .47 to .84 on factor 2 ($p$’s < .001). The two factors were significantly correlated ($r = .76, p < .001$), and factor 1 was determined to represent items relating to Harm Avoidance while factor 2 represented items relating to Incompleteness, supporting Summerfeldt’s (2004) hypotheses (Pietrefesa & Coles, 2008).

Summerfeldt (2004) proposed that incompleteness-based OCD would likely not respond as well as harm-avoidance-based OCD to traditional CBT, as the focus of CBT is changing interpretations of feared consequences, which play a minimal role at most in incompleteness-
based OCD. Summerfeldt (2004) also believed that feelings of incompleteness helped to explain why comorbidity patterns for OCD extended beyond traditional harm-avoidance anxiety disorders to include such repetitive disorders as Tic Disorder and Excoriation (skin picking) Disorder. In light of this proposed pattern of comorbidities, the recent reorganization of the DSM-5, which now places OCD in its own chapter along with disorders such as Excoriation and Trichotillomania (hair pulling), makes more theoretical sense (APA, 2013).

The last assertion Summerfeldt (2004) made regarding her motivational framework was that incompleteness was likely to be more important in the context of certain kinds of symptoms, such as symmetry, counting, and repetition behaviors, than others, such as harm-avoidant checking. These assertions have found partial support in more recent investigations. In a sample of 202 clients with OCD in Germany, researchers performed a hierarchal multiple regression on the two domains of the OC-TCDQ (Ecker & Gönner, 2008). In the first step of each regression, participant scores on measures of depression, anxiety, and obsessive-compulsive symptom severity were entered. After accounting for these measures of depression and anxiety, researchers found that both ordering and checking symptoms explained significant, unique variance on the Incompleteness subscale (Ecker & Gönner, 2008). Checking, along with obsessing, emerged as a significant predictor of harm avoidance. Most interestingly, in this investigation, obsessive-compulsive symptom severity showed higher correlations with Incompleteness than Harm Avoidance, though whether these differences were significant was not tested (Ecker & Gönner, 2008).

In Pietrefesa and Coles’s (2008) work, correlations between the two subscales of the OC-TCDQ and different symptoms of OCD were also examined. In their findings, Incompleteness
showed higher correlations than Harm Avoidance with the following symptoms: washing \( (r_{\text{harm}} = .35, r_{\text{inc}} = .45) \), checking \( (r_{\text{harm}} = .49, r_{\text{inc}} = .58) \), doubting \( (r_{\text{harm}} = .55, r_{\text{inc}} = .63) \), and ordering \( (r_{\text{harm}} = .41, r_{\text{inc}} = .59; \) Pietrefesa & Coles, 2008). Interestingly, these results and those of Ecker and Gönner (2008) contradict Summerfeldt’s (2004) expectation that checking will be related to harm avoidance but not highly related to incompleteness. Most important to this investigation, however, Incompleteness was more highly related than Harm Avoidance to ordering behaviors (Ecker & Gönner, 2008; Pietrefesa & Coles, 2008; Summerfeldt, 2004).

This pattern of relationships has been supported in behavioral studies as well as self-report studies. In a behavioral study to determine whether incompleteness versus harm avoidance would predict different behaviors in real situations, Pietrefesa and Coles (2009) devised a series of six tasks for undergraduate students to perform. Three tasks (washing dishes, sorting chemicals, and proofreading another student’s resume) were designed to elicit feelings of harm avoidance and three others (arranging books, hanging pictures, and sorting recycling) were designed to elicit feelings of incompleteness. Trait levels of Harm Avoidance and Incompleteness were also measured using the OC-TCDQ. Pietrefesa and Coles (2009) found that Harm Avoidance scores correlated significantly with anxiety in performing the behavioral tasks \( (r = .42, p < .01) \) as well as a desire to prevent harmful consequences of making mistakes on the tasks \( (r = .41, p < .01) \). Incompleteness scores correlated significantly with a desire to perform the tasks perfectly \( (r = .41, p < .01) \) as well as feelings of discomfort \( (r = .42, p < .01) \). Interestingly, Incompleteness was also related to a desire to prevent harmful consequences, but to a lesser extent than Harm Avoidance \( (r = .28, p < .01) \). Both Incompleteness \( (r = .32, p < .01) \)
and Harm Avoidance ($r = .34, p < .01$) were significantly correlated with hesitations, checks, and repetitions during the behavioral tasks (Pietrefesa & Coles, 2009), indicating that both are likely important to compulsive behaviors.

In an interesting study related to incompleteness-based OCD, Ferrao and colleagues (2012) confirmed that, instead of always being preceded by obsessions, some compulsions were preceded only by what they termed “sensory experiences.” In their sample of 1,001 patients diagnosed with OCD, they found that 65% endorsed some kind of sensory experience preceding or coinciding with their compulsions; the most common sensory experiences were “just right” experiences triggered by tactile, visual, auditory, or internal stimuli, as well as feelings of energy release or urges. Ferrao and colleagues (2012) found that age of onset of OCD symptoms was significantly younger among patients endorsing these sensory experiences, compared to those who did not (Pearson’s $\chi^2 = 12.1, p = .002$). Among OCD patients with sensory experiences, symmetry, ordering, and arranging symptoms were also significantly more common than in those without sensory experiences ($Yates\,\chi^2 = 78.3, p < .001$ [Yates $\chi^2$ is a corrected $\chi^2$ used in small samples]), and symmetry symptoms were the most common symptoms reported by the sensory experiences group (Ferrao et al., 2012). Incompleteness in the form of sensory experiences or urges does appear to be highly related to ordering concerns in OCD.

Thus, Summerfeldt (2004) and Summerfeldt and colleagues’ (2001) work represented an important extension of the focus of the cognitive-behavioral model, suggesting that, in addition to responsibility/harm avoidance, incompleteness should be considered.
The Importance of Symptom Type: Relationship to Belief Domains

Salkovskis’s (1985) cognitive-behavioral model was posited as an explanation for why symptoms arise and how they are maintained in people afflicted by OCD. However, subsequent research suggests that there may be important differences in the core cognitions or beliefs associated with different obsessions and compulsions. Underlying dysfunctional beliefs have been a subject of investigation in the OCD research literature for some time; as mentioned earlier, more than 20 years ago the OCCWG (1997) created a list of six belief domains hypothesized to be important to OCD: responsibility, overimportance of thoughts, importance of controlling thoughts, overestimation of threat, intolerance of uncertainty, and perfectionism. Years later, the OCCWG developed the Obsessive Beliefs Questionnaire (OBQ; OCCWG, 2003) to measure levels of these beliefs. The initial OBQ underwent revision and currently has a replicable three-factor structure, with the emerging factors being (a) Responsibility/Threat Estimation, (b) Perfectionism/Certainty, and (c) Importance/Control of Thoughts (OCCWG, 2005). This suggests that there are at least three main beliefs associated with OCD-related cognitions.

To examine whether these belief domains relate differentially to specific obsessive and compulsive symptoms, Julien, O’Connor, Aardema, and Todorov (2006) explored these beliefs in a study of 126 individuals diagnosed with OCD. They examined whether beliefs as rated on the OBQ would predict symptoms as measured by the Padua Inventory—Revised (PI-R; van Oppen, Hoekstra, & Emmelkamp, 1995), a measure of obsessive-compulsive symptoms. Their results revealed that, after accounting for scores on measures of depression and anxiety,
Responsibility/Threat Estimation significantly predicted rumination symptoms (adj. $R^2 = .54$, $p < .001$), Importance/Control of Thoughts significantly predicted impulse phobia symptoms (adj. $R^2 = .33$, $p = .002$), and Perfectionism/Certainty significantly predicted both checking (adj. $R^2 = .13$, $p = .002$) and precision (adj. $R^2 = .13$, $p = .019$) symptoms (Julien et al., 2006).

In a similar investigation, Taylor and colleagues (2010) used structural equation modeling to see which belief domains of the OBQ would predict which symptoms of another measure of obsessive-compulsive symptoms, the Obsessive-Compulsive Inventory—Revised (OCI-R; Foa et al., 2002). Taylor and colleagues (2010) sampled 5,000 undergraduates from universities across the United States and found that Responsibility/Threat Estimation significantly ($p < .01$) predicted all six OCI-R symptom subscales: Ordering, Checking, Neutralizing, Obsessing, Hoarding, and Washing. Importance/Control of Thoughts significantly predicted obsessing and, to a lesser extent, neutralizing and washing. Perfectionism/Certainty was found to significantly predict ordering (Taylor et al., 2010). Perfectionism/Certainty was also found to predict Ordering in a hierarchical regression on a sample of 99 patients with a primary diagnosis of OCD, after accounting for depression and trait anxiety scores in the first step ($\beta = .49$, $R^2 = .20$, $p < .05$; Tolin, Brady, & Hannah, 2008). These results, along with those of Julien and colleagues (2006), suggest that there may be different cognitive processes underlying different symptoms of OCD.

**Treatment and Classification**

Besides underlying cognitions, there is existing evidence of other clinically meaningful
differences among symptom domains. In this context, research examining the less explored domains of OCD (ordering, mental neutralizing, obsessions without compulsions) becomes an important practical consideration.

Some of the most notable differences involve differential responses to treatment. As mentioned above, symptoms which have been heavily researched show superior treatment response in therapy to those that have been less well researched (McKay et al., 2004; McLean et al., 2001), suggesting that therapeutic techniques which are effective for one type of symptom may not generalize to others. As an example, results from a randomized controlled trial of behavior therapy showed that the presence of sexual or religious obsessions significantly predicted poorer treatment outcomes after accounting for pretreatment symptom severity and depression ($R^2 = .17, p = .01$; Mataix-Cols, Marks, Greist, Kobak, & Baer, 2002). There are differences in pharmacological outcome studies as well, with ordering and symmetry concerns responding better to treatment with phenelzine (a monoamine oxidase inhibitor [MAOI]) than did other symptoms (Jenike, Baer, Minichiello, Rauch, & Buttolph, 1997).

Besides treatment, there appear to be other significant differences among symptoms. Eventually, due to its history of poorer treatment response, as well as statistical evidence that hoarding behavior seems less related to other OCD domains (Abramowitz et al., 2003; Wu & Watson, 2005), hoarding was reclassified as a disorder separate from OCD (APA, 2013). It is unlikely that all of the remaining OCD symptom domains will be deemed different enough to one another to merit their own separate disorder categories; however, as subtler differences between symptoms seem to exist, these differences should be explored in greater depth.
The Special Case of Ordering Concerns

Although ordering concerns are one domain of obsessive-compulsive symptoms that have received less research attention, there is some evidence to suggest that they may behave differently from the more “classic” symptoms of checking and contamination.

In an item-level exploratory factor analysis which Denys and colleagues (2004) conducted on the Y-BOCS Symptom Checklist scores of 335 patients with OCD, Symmetry emerged as one of five factors which accounted for significant variance. The results of this analysis do suggest that symmetry concerns are measurably distinct from contamination or checking concerns in patient samples. According to Denys and colleagues, a substantial proportion of their participants (57 of 214 participants with an identifiable primary symptom) indicated concerns loading on the Symmetry factor as their primary concern. The researchers found that, among their participants, those with primary symmetry concerns had a significantly lower age of onset than those with primary contamination concerns ($M_{\text{symmetry}} = 17; M_{\text{contamination}} = 23; p = 0.019$) and a significantly higher prevalence of OCD in first-degree relatives than all other groups (42% versus 19 – 35% for other groups, $p = 0.004$; Denys et al., 2004). These could indicate a higher degree of heritability for ordering concerns than other symptoms.

In a subscale-level factor analysis of the Y-BOCS Symptom Checklist scores of 169 participants diagnosed with OCD, Hasler and colleagues (2005) also found a Symmetry, Ordering, and Arranging factor which accounted for 18.6% of their overall variance. This Symmetry factor was one of four factors identified, along with an Aggressive/Sexual factor (accounting for 19.5% of the variance), a Contamination factor (13.7% of the variance), and a Hoarding factor (12.9% of the variance). Using logistic regression, they found that higher scores
on this Symmetry, Ordering, and Arranging factor were associated with significantly higher odds for being diagnosed with certain comorbid disorders, such as Bipolar Disorders ($OR = 1.5; 95\% CI: 1.1, 2.2; p < 0.05$), Panic Disorder ($OR = 1.7; 95\% CI: 1.3, 2.4; p < 0.001$), Agoraphobia ($OR = 2.2; 95\% CI: 1.5, 3.2; p < 0.001$), and Alcohol/Substance Use Disorders ($OR = 1.4; 95\% CI: 1.0, 1.9; p < 0.05$), whereas higher hoarding and contamination symptoms did not account for increased odds for suffering from any of these disorders (Hasler et al., 2005). Each of the four factors Hasler and colleagues (2005) identified showed different comorbidity patterns, further suggesting important clinical differences in obsessive-compulsive symptoms.

In a study that prospectively examined whether negative life events would predict future obsessive-compulsive symptoms, Coles and Horng (2006) found that, although negative life events were significantly related to total scores on an obsessive-compulsive symptom inventory at Time 2 ($r = .22, p < .001$), negative life events did not correlate significantly with ordering symptoms at Time 2, using their strict significance criterion ($r = .15, p > .001$ [n.s.]). Although their alpha level for these analyses was quite strict, ordering symptoms showed the second lowest correlation to prior negative life events of any of the symptoms measured by Coles and Horng (2006), second only to washing symptoms ($r = .07; p > .001$). This suggests again that ordering symptoms may function somewhat differently from other common OCD symptoms.

Returning to the Radomsky and Rachman (2004a) study of memory for organized versus disorganized objects in participants with subclinical ordering concerns, the researchers cited issues of personal relevance and responsibility as reasons why they were unable to attain significant findings. Although this explanation would be consistent with Salkovskis’s (1985) cognitive-behavioral model, there are other possible reasons why Radomsky and Rachman’s
(2004a) investigation of memory failed to find effects. Other possibilities listed include difficulty of the recall task or that other aspects of the manipulation were too distracting. One possibility that Radomsky and Rachman (2004a) did not explore was the possibility that ordering and arranging behaviors do not fit the cognitive-behavioral model as well as other symptoms. There is evidence that classic compulsions are performed in response to a specific, feared consequence (Rachman, 2002; Salkovskis, 1985), but perhaps ordering and arranging compulsions do not behave this way. To use Summerfeldt’s (2004) terminology, perhaps incompleteness, rather than harm avoidance, is important when considering ordering and arranging compulsions. As noted, there are certain differences among different types of obsessive-compulsive symptoms, so this would not be the first disparity to be found.

Other research by Radomsky and Rachman (2004b) provides some evidence that ordering and arranging compulsions are not performed due to specific feared consequences. In a study of college students with lower than average versus higher than average ordering concerns, participants with higher ordering concerns were interviewed about their thoughts and beliefs about disorganized environments (Radomsky & Rachman, 2004b). Of 24 participants with high ordering concerns, 22 (91.7%) said they arranged objects until they felt right. Only two participants identified more general feelings of impending doom as a reason for their arranging behavior, and even these two were unable to name any specific feared threats (Radomsky & Rachman, 2004b). These findings run counter to evidence that other compulsions, such as checking, are performed in response to specific feared events or consequences (Rachman, 2002) and provide additional support that perhaps incompleteness is more important to these kinds of symptoms (Ecker & Gönner, 2008; Pietrefesa & Coles, 2008; Summerfeldt, 2004).
The current study attempted to replicate Radomsky and Rachman’s (2004b) procedure to determine whether students who endorse higher levels of ordering concerns on questionnaires will (a) experience greater anxiety and/or urges to arrange items in a disorganized environment and (b) report incompleteness-based, as opposed to harm-avoidance-based, cognitions related to ordering and arranging behaviors. Since the original study may have suffered from relatively low power with a total of 48 participants across conditions (Radomsky & Rachman, 2004b), a larger sample was targeted. If replicated in a larger sample, findings such as these would indicate that individuals with ordering concerns struggle with feelings of anxiety when made to work in disorganized spaces. Such findings may provide further evidence for the importance of incompleteness to ordering symptoms and suggest that treatments based on considerations of harm avoidance may have limited efficacy in treating ordering symptoms (McKay et al., 2004; Sookman, Abramowitz, Calamari, Wilhelm, & McKay, 2005; Summerfeldt, 2004; Williams et al., 2013).

The Present Study

Three studies that examine ordering and arranging concerns are described in Radomsky and Rachman (2004b). The current study is a replication of Study 3.

In their original experiment, Radomsky and Rachman (2004b) drew a sample of participants $.5 SD$ above or below the mean score on a measure of ordering and arranging concerns. Like the original study, this replication also drew from a college student population. Although this project examined obsessive-compulsive phenomena in a subclinical sample,
previous research suggests that analogue samples can reveal information relevant to understanding clinical levels of these phenomena because clinical OCD appears to exist on a continuum with subclinical OCD (Abramowitz et al., 2014; Olatunji, Williams, Haslam, Abramowitz, & Tolin, 2008).

In the original study, baseline subjective units of distress scale (SUDS) ratings were gathered for all participants, and all participants then were told they would have five minutes to prepare a speech to present before three faculty members. Half of the participants prepared in a disorganized room and half prepared in an organized room, resulting in a 2x2 design (level of ordering concerns by room organization). After three minutes in the organized/disorganized room, participants gave SUDS ratings again, and those participants with high ordering concerns were briefly interviewed about ordering and arranging beliefs and cognitions (Radomsky & Rachman, 2004b). Information about these beliefs and cognitions was assessed through interview. Radomsky and Rachman’s (2004b) qualitative results suggested higher incompleteness issues than harm-avoidance issues in ordering behavior. If replicated, these findings may offer improved understanding as to why some individuals with OCD fail to improve through treatment or to maintain gains following exposure and response prevention training. In such resistant cases, individuals may be more influenced by incompleteness concerns than harm-avoidance concerns, indicating another, possibly more fruitful focus for therapy.

Hypotheses

Two major hypotheses, both consistent with Radomsky and Rachman (2004b), were
proposed for this project. Hypothesis 1 was that the high ordering concerns participants who prepared their speeches in the disorganized room would show significantly higher anxiety than each of the other three groups of participants (high ordering/organized room, low ordering/disorganized room, low ordering/organized room).

Hypothesis 2 was that participants in the high ordering concerns group would be unable to name a specific fear during the semi-structured interview. If found, such a result would be more consistent with a conceptualization of ordering behaviors that focuses on incompleteness rather than harm avoidance.
CHAPTER TWO

METHODS

Participants

Participants for the study were recruited from the Introductory Psychology subject pool at Northern Illinois University. Potential participants were screened through mass survey testing procedures at the beginning of each semester using the Schedule of Compulsions, Obsessions, and Pathological Impulses (SCOPI; Watson & Wu, 2005). Participants who met the inclusion criteria were invited via email to participate in exchange for partial course credit.

Sample Size

Multiple issues were considered when determining the targeted sample size for the current study. Radomsky and Rachman (2004b) did not report effect sizes, but an effect size of $d = .70$ was computed from the reported sample sizes and $F$ statistic. This $d$ was converted to an $f^2 = .12$, and a power analysis was conducted using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). Using the parameters of an alpha level of .05 and power of .80, it was estimated that the total sample size would need to be 95.

A targeted sample size of 100 was proposed for the current study; a final sample of 126 was used. This sample more than doubled the $N = 48$ from Radomsky and Rachman (2004b).
Measures

Demographic Questionnaire

To characterize the sample as to basic demographic variables, participants were presented with several questions regarding age, sex, and racial/ethnic identification. See Appendix A for the demographic questionnaire.

Obsessive-Compulsive Ordering Behavior

The Schedule of Compulsions, Obsessions, and Pathological Impulses (SCOPI; Watson & Wu, 2005) is a 47-item questionnaire that uses a 5-point response scale, ranging from 1 (strongly disagree) to 5 (strongly agree). It was developed from an item pool intended to cover a broad range of obsessive thoughts and compulsive behaviors as well as issues with impulse control, given the conceptual overlap between OCD and impulse control disorders (Watson & Wu, 2005). Factor analysis of the SCOPI reveals five subscales: Obsessive Checking (14 items), Obsessive Cleanliness (12 items), Compulsive Rituals (8 items), Hoarding (5 items), and Pathological Impulses (8 items). For the purposes of this study, only the Compulsive Rituals subscale (SCOPI-CR) was used to screen potential participants for high versus low ordering and arranging behavior.

The Compulsive Rituals subscale demonstrates good internal consistency (α = .88) and 2-month retest reliability (r = .82) among college students (Watson & Wu, 2005). It shows a strong correlation (r = .64; Watson & Wu, 2005) with the Ordering subscale of the Obsessive-
Compulsive Inventory – Revised (OCI-R; Foa et al., 2002) but was preferred due to its length (eight items of SCOPI Compulsive Rituals versus three items of OCI-R Ordering) and greater breadth of symptom content. Namely, the three OCI-R Ordering items are very similar (i.e., I get upset if objects are not arranged properly, I get upset if others change the way I have arranged things, and I need things to be arranged in a particular order). The SCOPI Compulsive Rituals subscale also shows lower correlations with measures of general distress, such as the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996) than does OCI-R Ordering with the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), suggesting it may have better discriminant validity (r = .04 for SCOPI Compulsive Rituals; r = .23 for OCI-R Ordering; Watson, 2009).

Another option would have been to use the Symmetry, Ordering, and Arranging Questionnaire (SOAQ) devised by Radomsky and Rachman (2004b) to measure ordering behaviors specifically. This measure has shown high internal consistency (α = .96) and retest reliability over a period of about 16 days (r = .92) in a sample of college students (Radomsky & Rachman, 2004b). However, similar to the OCI-R, the SOAQ mainly taps arranging content. Ultimately, the SCOPI-CR was chosen in consideration of both length (8 items vs. 20 in the SOAQ) and relative breadth. In the current study, the SCOPI-CR showed excellent internal consistency (α = .95). See Appendix B for the SCOPI-CR subscale.

Social Phobia

The Social Phobia Scale (SPS) and Social Interaction Anxiety Scale (SIAS; Mattick &
Clarke, 1998) are two complementary 20-item scales that were developed to assess circumscribed and generalized social fears, respectively. The SPS inquires about anxiety over being scrutinized in a number of specific social situations, whereas the SIAS inquires about anxiety in the context of conversation with others. Both scales use response options ranging from 0 (Not at all characteristic of me) to 4 (Extremely characteristic of me). In a student sample, short forms of the two scales were found to be correlated $r = .59$ (Fergus, Valentiner, McGrath, Gier-Lonsway, & Kim, 2012). The SIAS/SPS show sensitivity in discriminating between clinical groups with and without social phobia and in discriminating clinical samples from community and undergraduate samples (Mattick & Clarke, 1998).

Due to the length of these scales, and because they often are administered together, several groups of researchers have devised short forms of the SIAS/SPS (Carleton et al., 2009; Fergus et al., 2012; Peters, Sunderland, Andrews, Rapee, & Mattick, 2012). From among these, the SIAS/SPS short forms created by Fergus and colleagues (2012) were chosen for inclusion in this study. The SIAS/SPS short forms have several advantages over the original SIAS/SPS. First, the short forms only are six items each, compared to the original 20-item measures. This saved on time administering the scales to participants and possibly helped to avoid fatigue. Both short forms show adequate internal consistency for scales of that length (SIAS: $\alpha = .80$, SPS: $\alpha = .77$; Fergus et al., 2012). In addition, the short forms show strong correlations with their longer counterparts; both short forms correlated $r = .94$ ($p < .01$) with their respective original scale (Fergus et al., 2012). Lastly, these short forms are preferred over the full-length scales and other short forms because they were designed to be more easily understood by the general population; the SIAS short form has an estimated 6th-grade reading level, and the SPS short form has an estimated 5th-grade reading level (Fergus et al., 2012). In the current study, the SIAS showed
good internal consistency ($\alpha = 0.86$) and the SPS showed acceptable internal consistency ($\alpha = 0.79$). See Appendix C for a copy of the SIAS and SPS short forms.

**State Subjective Distress**

The Subjective Units of Distress Scale (SUDS; Wolpe, 1969) is a rating scale used by researchers and clinicians to determine momentary distress. It is a verbally administered, one-item “scale” ranging from 0 (absolute calmness) to 100 (the worst anxiety ever experienced). For the purposes of this investigation, baseline SUDS rating was used as a predictor variable in the regression analysis, and post-provocation SUDS rating served as the dependent variable.

**State Anxiety**

Because the SUDS measure consists only of one item, internal consistency cannot be examined. The State-Trait Inventory for Cognitive and Somatic Anxiety—State Version (STICSA-S; Ree, MacLeod, French, & Locke, 2000) also was administered at post-test. This provided a way to corroborate the findings and compensate for problems with the SUDS scale (e.g., face validity, differing interpretations among participants). The STICSA-S is a 21-item measure consisting of a 10-item Cognitive Anxiety subscale and an 11-item Somatic Anxiety subscale. Internal consistency of both subscales in a combined student and clinical sample was found to be good ($\alpha$s = .88), and the STICSA outperformed the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Luschene, 1970) in its ability to discriminate from depression.
symptoms (Grös, Simms, Antony, & McCabe, 2007). In the current study, the STICSA demonstrated excellent internal consistency ($\alpha = .96$). See Appendix D for a copy of the STICSAS.

**General Distress**

The Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) is a 20-item self-report questionnaire with response options ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). The PANAS consists of two 10-item scales: a Positive Affect (PA) subscale and a Negative Affect (NA) subscale. The instructions ask participants to endorse how often they have felt 10 positive moods and 10 negative moods and may be altered to account for time frames as short as the current moment or as long as “in general” (Watson et al., 1988). For the purposes of this study, participants were asked how often they have felt these moods in general, and only the negative affect items were used, as the goal of this measure was to examine trait levels of negative affectivity.

The “in general” instructions of the PANAS show good internal consistency for the NA subscale ($\alpha = .87$; Watson et al., 1988). In the same study, this subscale also demonstrated good retest reliability over an 8-week interval in college students ($r = .71$). The NA subscale showed the expected low correlation with the PA subscale ($r = -.17$), since the researchers were attempting to measure PA and NA as two separable, orthogonal constructs (Watson et al., 1988). The NA subscale distinguished between psychiatric inpatients and college students and showed appropriate correlations with existing measures of depression and anxiety; as expected, NA
correlates positively to measures of psychopathology (Watson et al., 1988). PANAS-NA can be used as a measure of general distress. In the current study, the PANAS-NA showed good internal consistency ($\alpha = .89$). See Appendix E for a copy of the PANAS-NA.

Qualitative Information About Ordering Behaviors

Qualitative information relating to ordering and arranging behaviors was gathered using the Ordering and Arranging Behaviors Interview (see Appendix F). The Ordering and Arranging Behaviors Interview consists of four questions relating to thoughts and motivations associated with urges to order or arrange surroundings. This interview was taken from Radomsky and Rachman (2004b).

Procedure

Selection

Participants were drawn from students enrolled in Psychology 102. Students were screened using the mass survey at the start of each semester. The SCOPI Compulsive Rituals subscale was administered at this time, and students scoring either $.5\ SD$ above or below the current sample’s mean scores were invited through SONA to participate in the study. The study used $.5\ SD$ above and below the mean to create “high” and “low” ordering concerns groups, respectively. Use of $.5\ SD$ is a more liberal criterion than the common practice of using $1\ SD$ to
assign “high” and “low” extreme groups; however, it affords the practical advantage of increasing the likelihood of recruiting a sufficient sample size and also precisely replicates Radomsky and Rachman (2004b).

The .5 SD cut-off score from the mass survey data for the first semester of data collection was used as the cut-off for each semester. This was done for two reasons: (1) to maintain consistency in study procedures and (2) because the means and SD of mass survey participants were not expected to vary meaningfully from one semester to the next, as the population primarily was first-year undergraduates at NIU each semester.

**Randomization**

Randomization of participants to either the “organized” or “disorganized” room conditions occurred before participants arrived to the study. The 50 participants in the high ordering concerns group were split into five blocks of 10; within each block of 10, a random numbers generator was used to allocate five participants to the organized and five participants to the disorganized conditions. The same procedure was used to allocate the low ordering concerns group. More randomized conditions were added, in blocks of 10, as needed. This method retained a high degree of randomness while preventing uneven allocation, such as a majority of high-ordering participants ending up in the organized office space or another problematic combination.
Baseline Measures

Those students who were invited to participate in the study were administered the SCOPI-CR subscale again through Qualtrics upon their arrival to participate, to ensure that they still met criteria for inclusion in either the “high” or “low” group. Those who had regressed toward the mean and whose scores fell between -.5 and .5 $SD$ (or 16 and 25) were retained for the duration of the study and given full credit. However, their data were not utilized in the primary analyses for this study. The SCOPI-CR was administered again at baseline, instead of another measure of ordering behaviors, in order to maximize confidence that the same construct was being measured.

The remaining eligible participants completed the PANAS-NA and the SIAS/SPS, via Qualtrics. Participants then were to provide a baseline SUDS rating, which was used as a covariate in the regression analysis, in order to conceptually replicate Radomsky and Rachman (2004b). After the baseline questionnaires were completed, participants were asked if they had a current or prior diagnosis of OCD; in order to minimize potential harm, those who reported a positive history ($n = 4$) were dismissed from the study at this point with full credit.

Manipulation

Next, all participants were asked to prepare a 5-minute speech to be given before two faculty members. Participants were told that, since this activity often produces a certain degree of stress, they would be given a private space in which to prepare. Participants then were directed to
an office space. Half of the participants prepared their speeches in an office which had been intentionally organized to be unusually neat, and half of the participants prepared their speeches in the same office but which had been intentionally disorganized to be unusually messy. Please see Appendix I for images of the disorganized office space and Appendix J for images of the organized office space.

**Dependent Variables**

After three minutes in the given office space, participants were asked for a SUDS rating again. This 3-minute time frame is a direct replication from Radomsky and Rachman’s (2004b) procedure. It was hypothesized that, since all participants were exposed to the speech manipulation, differences in post-provocation SUDS ratings between groups would reflect the effect of exposure to an organized versus disorganized room stimulus. As in Radomsky and Rachman’s (2000b) study, all participants, regardless of group or condition, were asked if they arranged items in the office or felt urges to arrange items in the office. If they reported that they did arrange items or felt urges to arrange items, the same short interview given by Radomsky and Rachman (2004b) was administered (see Appendix E). If participants indicated they did not arrange or have urges to arrange items in the office, the interview was skipped.

Beyond providing qualitative information, the question of whether participants arranged or felt urges to arrange items in the office also was used as a manipulation check. If the office stimulus was effective, it was expected that a high proportion of participants in the high ordering
concerns group/disorganized room condition would feel urges to arrange the space, whereas a relatively low proportion of participants in the other three conditions would feel those urges.

In addition to self-report, a picture of the office space was taken after each participant’s time in the space concluded. These post-provocation pictures were compared to pre-provocation pictures of the office space to determine if any items had been moved (regardless of whether participants reported having moved any).

**Debriefing**

Finally, participants were debriefed as to the true nature of the study, and the deception that was used to elicit anxiety in the office space situation. Participants were offered the opportunity to ask the researcher questions, were given a slip confirming their participation, and were thanked and dismissed. For a flow-chart representation of the full procedure, please refer to Appendix H.

**Replications and Extensions**

For the current study, some of the measures differed from Radomsky and Rachman’s (2004b) protocol: (1) This procedure used the SCOPI-CR (Watson & Wu, 2005) to sort participants into high and low groups, instead of the SOAQ (Radomsky & Rachman, 2004b). The SCOPI-CR was selected in place of the SOAQ due to its relative brevity, as well as item content which appears to assess a broader range of order-related behaviors than does the SOAQ.
(2) The baseline measure for general distress was changed from the BDI (Beck et al., 1961) to the PANAS-NA (Watson et al., 1988). Because the PANAS-NA measures negative affectivity, including terms such as “jittery” and “hostile,” it likely encompasses a broader range of negative affect than the BDI, which is depression focused. The PANAS-NA also is significantly shorter: 10 items as opposed to the 21 items of the BDI-II. (3) This procedure used the SIAS/SPS instead of the SPAI (Turner et al., 1989), as Radomsky and Rachman (2004b) used, as measures of social anxiety and as covariates in the regression analysis. The SIAS/SPS has advantages over the SPAI, namely, its far shorter length and more easily readable item content.

Much of the manipulation was a direct replication from Radomsky and Rachman (2004b). The room stimulus itself could not be replicated precisely. Attempts were made to gather information from the original authors about the organized and disorganized rooms, but they indicated that those records were not retained or available. The room stimulus for the present study was qualitatively examined using laboratory assistants to increase the likelihood that the organized and disorganized conditions were experienced as such. As with Radomsky and Rachman’s (2004b) protocol, SUDS ratings were procured prior to and after a 3-minute exposure to the organized or disorganized stimulus. Before being debriefed, this procedure made use of the same ordering and arranging interview that Radomsky and Rachman (2004b) used. However, to maximize the data being collected in the current study, the interview was administered to all participants, not just those in the high ordering groups.
CHAPTER 3
DATA ANALYSIS AND RESULTS

Data Cleaning

Before the main analyses were conducted, a systematic approach to data cleaning and preliminary evaluation was performed. First, individuals were excluded from the dataset if they showed invalid responses to either of the two validity questions imbedded within the questionnaires ($n = 2$). For the validity questions used, see Appendix G.

Second, patterns of missingness within the remaining dataset were analyzed using Little’s MCAR test. Little’s MCAR test revealed that the data were missing completely at random, $\chi^2(2121, N = 126) = 140.04, p = 1.00$. Participants with more than 5% of their data missing from a given questionnaire ($n = 1$) were excluded from analyses (Schafer, 1999). When less than 5% of data were missing from a given questionnaire, the missing values were computed using multiple imputation, a method by which multiple likely values are estimated from the dataset, and these likely values are averaged (Rubin, 1987). Based on guidelines established by Rubin (1987) and Schafer and Olsen (1998), ten imputations were used for this dataset. Eleven participants provided such missing values in the current dataset; 12 total data points were imputed.

The dataset next was examined for outliers by plotting all continuous measures using box plots. In addition, standardized total scores for all measured variables and subscales were examined. Absolute standardized values greater than 3.29 were deemed outliers (Field, 2005)
and winsorized to the nearest standard deviation, or an absolute standardized value of 3.00; three of these outlier values were identified. All cases identified as outliers by the box plots or standardized values were examined to ensure that the data were entered correctly. When the extreme score was confirmed to have been entered correctly (which was determined to be the case for all three identified outliers), the score was winsorized. Winsorizing an extreme score retains information in the dataset while decreasing the impact of a single extreme score (Ghosh & Vogt, 2012).

The normality of the independent variables was examined using the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. If either of these tests is significant, this indicates that the data are non-normal. Because these tests are sensitive to violations of normality, an alpha level of .01 was chosen a priori to indicate significant deviations from normality. Kolmogorov-Smirnov or Shapiro-Wilk tests were significant for the following variables: pre-provocation SUDS ratings, SPS scores, SIAS scores, and PANAS-NA scores. Please see Table 1 for full normality test results. Of note, the screening variable, SCOPI-CR, also was significantly non-normal, but this was expected; high and low groups were formed using this variable, resulting in a roughly bimodal distribution.

Table 1. Normality Tests for Predictor Variables

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov</th>
<th></th>
<th>Shapiro-Wilk</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>df</td>
<td>p</td>
<td>W</td>
</tr>
<tr>
<td>Pretest SUDS</td>
<td>.29</td>
<td>125</td>
<td>&lt;.001</td>
<td>.72</td>
</tr>
<tr>
<td>SCOPI-CR</td>
<td>.19</td>
<td>125</td>
<td>&lt;.001</td>
<td>.91</td>
</tr>
<tr>
<td>SIAS</td>
<td>.10</td>
<td>125</td>
<td>.005</td>
<td>.96</td>
</tr>
<tr>
<td>SPS</td>
<td>.15</td>
<td>125</td>
<td>&lt;.001</td>
<td>.90</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>.16</td>
<td>125</td>
<td>&lt;.001</td>
<td>.87</td>
</tr>
</tbody>
</table>
Next, the skew and kurtosis of the data were examined. The data were considered significantly skewed and/or kurtotic if standardized scores for skew and kurtosis exceeded 2.58 (Field, 2005). The following variables were found to be significantly skewed and/or kurtotic: pre-provocation SUDS ratings, SPS scores and PANAS-NA scores. For these skewed measures, square root or logarithmic transformations were considered. Logarithmic transformations appeared to be most useful in normalizing the skewed variables, and so these measures (pre-provocation SUDS, SPS, PANAS-NA) were used in their logarithmic-transformed state for the main analyses. Please see Table 2 for full descriptive statistics for independent variables.

Table 2. Independent Variable Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>Skew</th>
<th>S.E. Skew</th>
<th>Kurtosis</th>
<th>S.E. Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest SUDS*</td>
<td>2.48</td>
<td>1.23</td>
<td>-.28</td>
<td>.25</td>
<td>-.66</td>
<td>.50</td>
</tr>
<tr>
<td>SCOPI-CR</td>
<td>21.46</td>
<td>9.04</td>
<td>.03</td>
<td>.22</td>
<td>-1.40</td>
<td>.43</td>
</tr>
<tr>
<td>SIAS</td>
<td>14.70</td>
<td>5.29</td>
<td>.41</td>
<td>.22</td>
<td>-.49</td>
<td>.43</td>
</tr>
<tr>
<td>SPS*</td>
<td>2.41</td>
<td>.37</td>
<td>.69</td>
<td>.22</td>
<td>-.26</td>
<td>.43</td>
</tr>
<tr>
<td>PANAS-NA*</td>
<td>2.87</td>
<td>.79</td>
<td>.93</td>
<td>.22</td>
<td>.30</td>
<td>.43</td>
</tr>
</tbody>
</table>

*Note: * = log-transformed variable

The data also were tested for homogeneity of variance among all four groups on all variables to be examined. Levene’s test was used to assess for equal variance. Levene’s test suggested that none of the transformed predictor variables demonstrated significant heterogeneity of variance (ps ranged from .113 - .708).

Preliminary Analyses

Descriptive information for the full data set was gathered, including sex, age, and racial
identification. From a total of 211 participants proctored, 126 met criteria for inclusion in the main analyses and comprise the final sample (please see Appendix K for reasons for data loss in the current study). These participants were 64.3% female and averaged 19.80 years of age ($SD = 2.45$). The racial breakdown of the sample was 64.3% European American, 9.5% African American, 10.3% Asian/Asian American, 15.9% multiracial or “other” race. The majority of participants self-identified their ethnicity as non-Hispanic/Latino (84.9%). Overall, this sample is similar to others collected in this laboratory using a similar method (e.g., Conley & Wu, in press) and is considered to be representative of the broad undergraduate population of NIU.

Internal consistency and average inter-item correlations (AICs) were examined for all measures across all groups to check that measures performed with expected reliability across the sample. Low reliability of measures in any of the groups would be considered a study limitation, as low reliability of measures limits the degree of internal validity of a study. As mentioned in the measures section, SCOPI-CR scores, SIAS scores, PANAS-NA scores, and STICSA scores showed internal consistency that was good or better. SPS scores demonstrated acceptable internal consistency ($\alpha = .79$). The average inter-item correlations (AICs) were acceptable for most of the measures used in this study; however, the AIC for the SCOPI-CR was somewhat high (.69), indicating that the construct measured by this scale may be more narrowly defined than was expected. For full AIC information, please see Table 3.

<table>
<thead>
<tr>
<th>Measure Name</th>
<th>$\alpha$</th>
<th>AIC</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOPI-CR</td>
<td>.95</td>
<td>.69</td>
<td>8</td>
</tr>
<tr>
<td>SIAS</td>
<td>.86</td>
<td>.51</td>
<td>6</td>
</tr>
<tr>
<td>SPS</td>
<td>.79</td>
<td>.39</td>
<td>6</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>.89</td>
<td>.46</td>
<td>10</td>
</tr>
<tr>
<td>STICSA</td>
<td>.96</td>
<td>.53</td>
<td>21</td>
</tr>
</tbody>
</table>
Confounding Variables Test

A two-tailed $t$ test was used to compare baseline SUDS ratings between high and low ordering concerns groups. Because these two groups differ in levels of one kind of anxiety-relevant symptom (ordering concerns), there was a reasonable likelihood they would differ in others. As it is, the groups did not differ significantly on baseline SUDS ($t(124) = -1.78, p = .078$). Therefore, it was not deemed necessary to include baseline SUDS as a predictor variable in the main analyses.

A two-tailed $t$ test also was used to compare PANAS-NA scores between the high and low ordering concerns groups, testing again whether the high-ordering participants reported significantly higher trait negative affect than the low-ordering participants. Because “high” and “low” groups were created based on self-reported ordering and arranging concerns—a relatively narrow construct—it was not certain that there would be baseline group differences on the broad construct of negative affect. However, the mean-level difference was significant ($t(124) = -2.68, p = .008$), so PANAS-NA scores were accounted for in the main analyses.

Primary Analyses

Hypothesis 1

Hypothesis 1 stated that individuals in the high ordering concerns group who prepared their speeches in a disorganized room would show significantly higher anxiety than all other groups—high and low ordering concerns individuals who prepared their speeches in an
organized room as well as low ordering concerns individuals who prepared their speeches in a disorganized room. Hypothesis 1 was tested using multiple regression. As with Radomsky and Rachman’s (2004b) original study, baseline SIAS/SPS ratings were included as covariates in order to be certain the results were not solely the result of social anxiety due to the prospect of giving a speech. Because PANAS-NA scores were significantly different between groups, this measure was included as a covariate in the main analyses. Room condition and level of ordering concerns were entered as dummy-coded variables. The interaction of room condition by level of ordering concerns also was included. Post-provocation SUDS rating served as the dependent variable. This regression and all follow-up regressions were modeled as moderations in the PROCESS macro using Model 1 and 1,000 bootstrapping samples (Hayes, 2013).

Inconsistent with Hypothesis 1, the group with the highest post-provocation SUDS rating was the high ordering concerns/organized room group ($M = 46.28, SD = 5.11$), as opposed to the high ordering concerns/disorganized room group ($M = 36.65, SD = 4.95$). The full model contributed significantly to the prediction of post-provocation SUDS ratings ($R = .56, p < .001$). However, the interaction term (room condition x ordering concerns) did not significantly increase the predictive power of the regression, $\Delta R^2 = .00, p = .494$. Therefore, Hypothesis 1 was not supported in the full sample. For group means and full regression results using SUDS ratings as the outcome variable in the full sample, please see Table 4, Table 5 and Table 6.

Table 4. Group Means on SUDS, Full Sample

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>36.65</td>
</tr>
<tr>
<td>Control</td>
<td>46.28</td>
</tr>
</tbody>
</table>
Table 5. Model Summary, Full Sample, SUDS as Outcome

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$R^2$</th>
<th>$F$</th>
<th>df1</th>
<th>df2</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.56</td>
<td>.32</td>
<td>9.15</td>
<td>6</td>
<td>119</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 6. Regression Results, Full Sample, SUDS as Outcome

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>S.E.</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-102.06</td>
<td>25.22</td>
<td>-4.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Condition</td>
<td>-2.19</td>
<td>6.45</td>
<td>- .34</td>
<td>.734</td>
</tr>
<tr>
<td>Group</td>
<td>6.65</td>
<td>6.47</td>
<td>1.03</td>
<td>.306</td>
</tr>
<tr>
<td>Condition*Group</td>
<td>-6.10</td>
<td>8.88</td>
<td>- .69</td>
<td>.494</td>
</tr>
<tr>
<td>SPS</td>
<td>22.91</td>
<td>8.80</td>
<td>2.60</td>
<td>.010</td>
</tr>
<tr>
<td>SIAS</td>
<td>- .39</td>
<td>.76</td>
<td>.61</td>
<td>.613</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>30.64</td>
<td>3.23</td>
<td>.00</td>
<td>.002</td>
</tr>
</tbody>
</table>

These results—in particular, the mean SUDS ratings—were surprising, as Radomsky and Rachman (2004b) found the opposite pattern in their original study. Because of this unexpected pattern, the regression analysis was repeated among only more extreme participants (i.e., those who were at least ± 1 $SD$ from the mean on SCOPIC Compulsive Rituals) to examine whether Hypothesis 1 would hold among a subsample in which the high and low groups would be more different from one another. In this smaller sample ($N = 67$), the group with the highest post-provocation SUDS rating was in fact the high ordering concerns/disorganized room group ($M = 47.77$, $SD = 6.76$), as originally expected; however, the high ordering concerns/organized room group did not report appreciably lower ratings ($M = 45.38$, $SD = 7.00$). The full model contributed significantly to the prediction of post-provocation SUDS ratings ($R = .50$, $p = .007$). In this subsample, the interaction term did not significantly increase the predictive power of the regression equation, $\Delta R^2 = .00$, $p = .929$. Therefore, even among the more extreme subsample, Hypothesis 1 only was partially supported. For group means and full regression results using SUDS ratings as the outcome in the extreme sample, see Table 7, Table 8 and Table 9.
Table 7. Group Means on SUDS, Extreme Sample

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Experimental</td>
<td>47.77</td>
</tr>
<tr>
<td>Control</td>
<td>45.38</td>
</tr>
</tbody>
</table>

Table 8. Model Summary, Extreme Sample, SUDS as Outcome

<table>
<thead>
<tr>
<th>R</th>
<th>R²</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>.25</td>
<td>3.28</td>
<td>6</td>
<td>60</td>
<td>.007</td>
</tr>
</tbody>
</table>

Table 9. Regression Results, Extreme Sample, SUDS as Outcome

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-86.09</td>
<td>37.36</td>
<td>-2.30</td>
<td>.023</td>
</tr>
<tr>
<td>Condition</td>
<td>3.74</td>
<td>9.03</td>
<td>.41</td>
<td>.680</td>
</tr>
<tr>
<td>Group</td>
<td>11.23</td>
<td>9.77</td>
<td>1.16</td>
<td>.253</td>
</tr>
<tr>
<td>Condition*Group</td>
<td>-1.15</td>
<td>12.85</td>
<td>-0.09</td>
<td>.929</td>
</tr>
<tr>
<td>SPS</td>
<td>24.18</td>
<td>12.58</td>
<td>1.92</td>
<td>.059</td>
</tr>
<tr>
<td>SIAS</td>
<td>-1.17</td>
<td>1.11</td>
<td>-1.05</td>
<td>.297</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>25.94</td>
<td>14.09</td>
<td>1.84</td>
<td>.071</td>
</tr>
</tbody>
</table>

The regression analysis was repeated again using the STICSA-S, to rule out the possibility that the non-replication of Radomsky and Rachman’s (2004b) original findings was due to problems with the SUDS as the outcome measure (i.e., face validity, inability to test internal consistency, differing interpretations of the meaning of the 0 and 100 anchor points). For this regression analysis, the full sample (±0.5 SD from the mean) was used. Inconsistent with Hypothesis 1, the group with the highest post-provocation STICSA-S ratings again was the high ordering concerns/organized room group (M = 43.38, SD = 2.63) as opposed to the high ordering concerns/disorganized room group (M = 38.20, SD = 2.55). The full model contributed significantly to the prediction of post-provocation STICSA-S ratings (R = .68, p < .001). The interaction term (room condition x ordering concerns) did not significantly increase the predictive power of the regression, although it did approach significance, ΔR² = .01, p = .077.
Therefore, Hypothesis 1 was not supported in the full sample when the STICSA-S was used as the dependent variable in place of the SUDS. For group means and full regression results using STICSA-S ratings as the outcome in the full sample, please see Table 10, Table 11 and Table 12.

Table 10. Group Means, STICSA-S, Full Sample

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>38.21</td>
<td>35.50</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>43.38</td>
<td>33.24</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Model Summary, Full Sample, STICSA-S as Outcome

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$R^2$</th>
<th>$F$</th>
<th>df1</th>
<th>df2</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.68</td>
<td>.46</td>
<td>16.70</td>
<td>6</td>
<td>118</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 12. Regression Results, Full Sample, STICSA-S as Outcome

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>S.E.</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-45.79</td>
<td>11.78</td>
<td>-3.89</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Condition</td>
<td>3.00</td>
<td>3.01</td>
<td>1.00</td>
<td>.320</td>
</tr>
<tr>
<td>Group</td>
<td>6.18</td>
<td>3.01</td>
<td>2.05</td>
<td>.042</td>
</tr>
<tr>
<td>Condition*Group</td>
<td>-7.35</td>
<td>4.12</td>
<td>-1.78</td>
<td>.077</td>
</tr>
<tr>
<td>SPS</td>
<td>11.24</td>
<td>4.07</td>
<td>2.76</td>
<td>.007</td>
</tr>
<tr>
<td>SIAS</td>
<td>.12</td>
<td>.35</td>
<td>.35</td>
<td>.729</td>
</tr>
<tr>
<td>PANAS-NA</td>
<td>7.85</td>
<td>4.41</td>
<td>4.05</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Hypothesis 2

Hypothesis 2 stated that participants who arrange their surroundings or feel urges to arrange their surroundings would be unable to name a specific fear causing them to do so during the qualitative interview. Hypothesis 2 was tested through an examination of qualitative responses to the interview items. In the dataset, answers to interview items were coded either as indicating no specific feared consequences (0) or indicating a specific feared consequence (1). The frequency of these dummy-coded values was examined to see whether the ability to name a
specific feared consequence was infrequent among participants with urges to order or arrange. Approximately 8% of Radomsky and Rachman’s (2004b) participants endorsed a specific fear, and this was considered infrequent in their published report. Although somewhat arbitrary beyond its status as a published empirical result, the same bar was used to interpret the results of the proposed study.

In the full sample, 26 participants (20.6%) endorsed urges to order or arrange. Interestingly, although the majority of participants reporting urges to order or arrange were part of the high ordering concerns/disorganized room condition ($n = 14$), a sizeable minority were part of the low ordering concerns/disorganized room condition ($n = 9$). The remaining three participants were part of the high ordering concerns/organized room condition. Of the 26 total participants who had urges to order or arrange, the majority of responses to Question 3 (“What do you think would happen if you were unable to order or arrange your surroundings?”) did not indicate a specific fear (e.g., “nothing” or “don’t know”); however, six participants reported specific feared consequences (23%). These fears usually revolved around the speech task (e.g., won’t be able to focus, will slow me down on my speech). Therefore, Hypothesis 2 is not considered supported per the a priori bar for considering specific fears to be infrequent. The status of ordering symptoms as being more related to incompleteness concerns than to harm-avoidance concerns remains in question after examination in this relatively small, analogue sample.
CHAPTER FOUR
DISCUSSION

The purpose of the current study was to attempt to replicate the findings of Radomsky and Rachman (2004b) in a larger sample and with different self-report measures of the constructs under study, with the aim of increasing power and examining for evidence of generalizability and robustness of findings, respectively.

As in Radomsky and Rachman’s (2004b) original investigation, it was hypothesized that participants with high levels of ordering concerns would be more anxious after preparing a speech in a disorganized office space than would participants with high levels of ordering concerns in an organized office space and participants with low levels of ordering concerns in either type of office space. This hypothesis was partially supported. On one hand, among the full sample of individuals, the pattern was the reverse of what was expected; that is, participants with high ordering concerns in the organized office space reported the greatest anxious distress at post-test. Conversely, when the sample was restricted to only those most extreme participants, the results were as hypothesized. However, it must be noted that the high ordering concerns/disorganized participants were only modestly more anxious than were the high ordering concerns/organized participants. When the STICSA was used as the outcome measure in the full sample, the results also did not support the hypothesis. Participants with high ordering concerns in the organized office space again reported the greatest anxiety at post-test.
Interestingly, although all three regressions significantly predicted changes in the outcome variables (SUDS or STICSA-S), none of the analyses suggested that either room condition or group membership (or their interaction) significantly predicted the differences in these anxiety scores at post-test. Because the analyses controlled for participants’ levels of social anxiety and baseline negative affect, it is difficult to say what variable may have accounted for these differences. It likely was not one of the variables included in the present analysis; as such, these findings are difficult to explain with the model used in this study.

The second hypothesis also was not supported. Of participants who reported urges to order or arrange, 23% of them were able to identify a specific feared consequence. In keeping with Radomsky and Rachman (2004b), the cut-off for Hypothesis 2 to be considered supported was 8%. Although this stringent cut-off was not met, those participants who were able to cite a specific reason for their ordering behaviors were in the minority. Based on the results for Hypothesis 2, it is difficult to speak conclusively regarding the aforementioned assertion that incompleteness may be a more important predictor of ordering and arranging behaviors than harm avoidance (Ecker & Gönner, 2008; Pietrefesa & Coles, 2008; Summerfeldt, 2004). If harm avoidance were motivating ordering and arranging behavior, it is likely that the participants would have known more clearly what harm they wished to avoid (Salkovskis, 1985, 1999); only a minority of participants did. However, as the number was a non-zero value and in fact a significant minority, this issue will require further investigation. Were the number of participants around 5% or less, one could have greater confidence that incompleteness as opposed to harm avoidance was motivating these behaviors.
Limitations

Cronbach and Meehl (1955) suggested there are three possible reasons why studies generate null results: (1) the measures used were incorrect measures of the variables of interest; (2) the theory which generated the hypothesis is incorrect; or (3) the experiment was an incorrect test of the hypothesis. In the current study, it seems least likely that (1) is the reason for the null findings. All baseline measures used have been examined and validated extensively elsewhere and demonstrate internal consistency estimates that are good or better in the current sample (the SPS being the one exception, which demonstrated acceptable internal consistency per Cronbach’s alpha and good internal consistency per the AIC). The measures initially were chosen for their content validity (e.g., the SIAS contains items such as, “I am tense mixing in a group”; SUDS prompts “what is your current anxiety?”). Thus, the researcher has confidence that most of the measures used are tapping the constructs of interest in an accurate manner.

Despite this confidence in the other measures used in this study, some issues arose regarding SUDS ratings, specifically. First, this rating system has such high face validity that there is a possibility that some participants were able to infer its purpose. As such, demand characteristics may have affected these findings. However, it is hoped that by limiting SUDS ratings to only pre- and post-provocation SUDS, there was less likelihood of demand characteristics than if SUDS ratings had been assessed at multiple times across the organized/disorganized room manipulations. It also is hoped that these demand characteristics cancelled out across participants (e.g., some participants trying to present as more distressed to “help” the researcher, others trying to present as less distressed for social desirability purposes).
In future studies, this issue may be addressed through the use of psychophysiological measures, such as blood pressure or heart rate monitors, which generally are less susceptible to intentional or unintentional bias as compared to verbal self-reports.

The SUDS ratings provided an additional limitation to this study. Participants seemed to interpret the 0-100 response scale in a variety of ways. For example, a SUDS rating of 50 did not mean the same thing to all participants. In some cases, participants showed signs of underreporting (e.g., reporting SUDS below 30 despite displaying jitteriness and a fearful face); in others, participants potentially overreported (e.g., reporting a baseline SUDS of 70). It is my belief that these differences in scale interpretation would have evened out among participants across groups, such that they did not affect the main study findings. However, the issue does raise questions as to the validity of these data. It was for this reason that follow-up analyses using the STICSA were conducted. It is recommended that future research in this area make use of multiple self-report measures of state anxiety, as this study did reach slightly different conclusions based on which outcome measure (STICSA-S or SUDS) was used.

The second possibility is that the theory which generated the hypothesis is incorrect. In considering the theory, the following limitations are noted, with suggestions for changes to the theory and consequently the experimental design in future research attempts.

In its narrowest form, the theory which generated the hypotheses is that self-report measures correlate with actual behavior; in this case, questionnaire-based ordering and arranging concerns correlate with ordering and arranging behavior in the laboratory. This, in essence, is the model underlying all research involving self-report measures. However, this model likely requires modification so as to account for other factors that are known to influence behavior,
such as attention, idiosyncratic stimulus value, or social desirability. For instance, inattention to the study stimuli may have attenuated the link between self-reported ordering concerns and actual ordering behavior. That is, participants may have attended to the speech task to the exclusion of the office space. This was not accounted for in the current experiment, but a future study could ask participants simple questions about the office space as a way of testing their attention to it. This change to the experiment would reflect a slight change in the theory: Self-reported ordering and arranging concerns correlate with ordering and arranging behavior when the person notices a disorganized environment.

Next, stimulus value could have affected the relationship between self-report and behavior. Participants—even those who reported high levels of ordering concerns in general—may not have felt sufficient motivation to organize a room that contained someone else’s objects or was not disorganized because of them; that is, there may have been issues with the salience of the stimulus. This may be the reason why room condition did not significantly predict post-provocation SUDS scores in the primary analysis. As Radomsky and Rachman (2004a) found in their unsuccessful manipulation of memory for disorderly objects, stimuli in studies of obsessive-compulsive tendencies must be salient to participants in order to work properly. Whereas other constructs within the field of OCD research have been successfully studied using standardized laboratory stimuli (e.g., checking behaviors [Arntz et al., 2007] and “not just right experiences” [Pietrefesa & Coles, 2009]), perhaps there is something unique about ordering and arranging behaviors that makes them more difficult to reproduce using standardized stimuli in a laboratory. This may indicate that future research using idiographic stimuli, such as participants’ own belongings, or a choice among several situations of the most likely to cause distress would
be more successful in eliciting anxiety. Perhaps in the case of ordering behaviors, general stimuli work best for those with greater sensitivity to the OC concerns under study (and could be why the full sample of participants responded differently than the extreme sample in the current study). Researchers have found that idiographic stimuli, such as those in the participants’ own homes, may have greater salience to participants (Lopatka & Rachman, 1995). Alternatively, if idiographic stimuli are not possible (e.g., in a university-based laboratory setting with limited space and/or resources), then other ways of increasing participants’ ownership of the disorganized space might be used to increase the salience of the stimulus. For example, have participants disorganize the space so that the state of disorder is “their fault” (though this approach may allow participants to “blame” the proctor as responsible) or make participants responsible for organizing it again (similar to the sorting task participants completed in Arntz et al., 2007). Lack of stimulus salience could have attenuated the relationship between report of ordering/arranging concerns and behavior in this study; therefore, changes such as these to the experimental design also may suggest a necessary change to the wording of the theory. Perhaps it would be more accurate to say that self-report of ordering and arranging behaviors predicts ordering behaviors under the condition of a sufficiently salient stimulus.

Finally, there is a chance that social desirability may have caused some participants to suppress their urges to arrange the objects, to the point of not wanting to report these urges to the researcher. One very basic approach to this issue in future studies may be inclusion of a measure of social desirability among the baseline questionnaires. Using this additional check, a future study operating under the theory that self-report of ordering and arranging behavior correlates
with actual arranging behavior if the participant is not suppressing/mis-reporting behavior due to social desirability may find better success.

The next two limitations are purely experimental in nature; that is, they represent two ways in which the experiment may have been an incorrect test of the hypothesis. First, a dosage problem may have occurred in the current study. Perhaps 3 minutes in a disorganized space was not enough to markedly change anxiety levels for any participants save those with the highest levels of vulnerability. This could explain both (a) why participants in the full sample were not most anxious in the high ordering/disorganized condition and (b) why room condition was not a significant predictor of anxiety scores at post-test. The findings of Radomsky and Rachman (2004b) suggested that 3 minutes would be sufficient, and since the proposed study was meant to replicate theirs, the decision was made a priori to adhere to their methods wherever possible. However, due to the mixed pattern of results, it is recommended that a follow-up study should seek to increase the dose of distress involved, such as by increasing the amount of time spent in the dis/organized space.

Finally, this study may suffer from a limitation noted by Gönner and colleagues (2009), that is, limiting the experimental investigation to only one symptom domain. Although this practice is common in research and this study was attempting to replicate the original (Radomsky & Rachman, 2004b) in this respect, some information may have been lost by using this method and ignoring other types of concerns which may have contributed to behavior in the disorganized room space (e.g., contamination or checking-based concerns). It is difficult to say how the results may have looked different if participants had been sorted into high and low groups based on symptom total scores as opposed to Compulsive Rituals subscale scores, but perhaps creating
groups of participants who generally were high OC or low OC would have created groups that were more different from one another and the findings would have supported the hypotheses. Given the overlap between ordering and other OC concerns, future research may find it worthwhile to examine these symptom domains together, as opposed to in isolation from each other.

Implications

Despite these limitations, this study has some broader implications for OCD research. One of the findings from the present study is that self-reports of obsessive-compulsive concerns only sometimes map onto actual behavior as performed/reported in the lab. That is, individuals who are preoccupied with order may find it more difficult to work within disorganized environments, but the findings were mixed on this point and other variables which were not accounted for in the current study may intervene between those self-reports and behavior. This finding, however, appears to be true for individuals with more extreme levels of ordering concerns in the current sample, and so it is likely that people with clinical levels of these symptoms would find working in such an environment to be stressful and unproductive. It is interesting that these results appeared despite the brief time participants spent in these surroundings and the fact that the disorganized objects did not belong to them—that is, they could take no responsibility for the state of things. Thus, individuals with ordering-specific OCD symptoms may struggle in a work or school environment that is perpetually disorderly, and this may add to the distress or impairment they experience as a result of their condition. Anxiety felt
in disorganized spaces such as at school or work may be part of the reason why recent research suggests that individuals with symmetry and ordering as their primary symptoms of OCD experience longer illness duration and greater symptom severity (Lochner et al., 2016).

Although the stringent criteria regarding incompleteness versus harm-avoidance reporting was not met in the current study, the current findings add to a literature which suggests that ordering symptoms within OCD may emerge due to feelings of incompleteness, rather than fear of and attempts to avoid harm. Because exposure and response prevention—the current gold standard treatment for OCD—is based on Salkovskis’s (1985) cognitive model of OCD that emphasizes harm avoidance, this finding may partially explain why ordering symptoms show poorer treatment response than other OCD symptoms in some studies (McLean et al., 2001; Williams et al., 2013). If enough research is gathered in support of this assertion, a different form of treatment—one meant to target feelings of incompleteness specifically—may be required for this subpopulation.

The fact that the current findings only partially replicated the original findings (Radomsky & Rachman, 2004b) suggests certain implications for behavioral research and replication more broadly. These findings point to the need for continued replication efforts in the behavioral sciences to confirm the generalizability of findings both across different groups of participants and different measures of the constructs of interest.

Finally, this study points to the ongoing issue in social sciences research of trying to map self-report measures onto actual behaviors. In the current study, being sorted into a high versus low group based on self-reported ordering concerns did not appear to correlate with actual ordering behavior. It is likely that there are intervening variables that were not considered in the
current study, such as motivation, or aspects of the social situation in which the experiment took place. This points to two future directions for behavioral research. First, researchers need to include measures of behavior in their studies, rather than relying exclusively on self-report. Second, when researchers have confidence that their measures of a given construct are valid, they need to search for other reasons why those measures may not map onto behavior. Clinical researchers need to consider constructs from other branches of psychology, such as social or cognitive psychology, to understand the gap between what participants report they do and what they actually do.
REFERENCES


Gönner, S., Ecker, W., & Leonhart, R. (2009). Diagnostic discrimination of patients with different OCD main symptom domains from each other and from anxious and depressive controls. Journal of Psychopathology and Behavioral Assessment, 31(3), 159-167. doi: 10.1007/s10862-008-9114-0


APPENDIX A

DEMOGRAPHIC INFORMATION QUESTIONNAIRE
1. What is your age? ________

2. What is your sex?
   1=Male
   2=Female
   3=Other

3. What is your racial identification?
   1=African American/Black
   2=Asian/Asian American
   3=Caucasian/European American
   4=Native American
   5=Multiracial
   6=Other

4. What is your ethnicity?
   1=Hispanic/Latino
   2=Not Hispanic/Latino

5. Have you ever been diagnosed with a mental illness?
   1=Yes
   2=No

   If yes, what was the diagnosis? __________________________

6. Are you currently taking any psychiatric medications?
   1=Yes
   2=No

   If yes, what medication(s)? __________________________

7. Are you currently seeking therapy for any reason?
   1=Yes
   2=No

   If yes, for what reasons?
APPENDIX B

THE SCHEDULE FOR COMPULSIONS, OBSESSIONS, AND PATHOLOGICAL IMPULSES – COMPULSIVE RITUALS SUBSCALE (SCOPI-CR)
1 = strongly disagree; the statement is definitely false
2 = disagree; the statement is mostly false
3 = neutral or cannot decide; the statement is about equally true and false
4 = agree; the statement is mostly true
5 = strongly agree; the statement is definitely true

___ 1. I like to follow a particular order as I dress myself each day.
___ 2. I often follow the same, fixed order in performing everyday tasks.
___ 3. I sometimes find myself rearranging things to make sure that everything is in the proper order.
___ 4. I have a number of different rituals (e.g., sorting or touching things in a particular way) that I follow in my everyday life.
___ 5. I like to do things in a particular order when I am getting ready for bed.
___ 6. I have little rituals that I follow even though I know they are silly.
___ 7. If I don’t do certain tasks in a specific order, I feel uncomfortable.
___ 8. There are certain routine tasks that I always perform in exactly the same way.
APPENDIX C

THE SOCIAL INTERACTION ANXIETY SCALE AND THE SOCIAL PHOBIA SCALE
The Social Interaction Anxiety Scale

Please review each statement, and indicate the degree to which you feel the statement is characteristic or true of you.

0 = Not at all characteristic or true of me
1 = Slightly characteristic or true of me
2 = Moderately characteristic or true of me
3 = Very characteristic or true of me
4 = Extremely characteristic or true of me

1. I become tense if I have to talk about myself or my feelings.
2. I tense-up if I meet an acquaintance in the street.
3. I feel tense if I am alone with just one other person.
4. I am nervous mixing with people I don’t know well.
5. When mixing in a group I find myself worrying I will be ignored.
6. I am tense mixing in a group.

The Social Phobia Scale

Please review each statement, and indicate the degree to which you feel the statement is characteristic or true of you.

0 = Not at all characteristic or true of me
1 = Slightly characteristic or true of me
2 = Moderately characteristic or true of me
3 = Very characteristic or true of me
4 = Extremely characteristic or true of me

1. I get nervous that people are staring at me as I walk down the street.
2. I fear I may blush when I am with others.
3. I would get tense if I had to sit facing other people on a bus or a train.
4. It would make me feel self-conscious to eat in front of a stranger at a restaurant.
5. I get tense when I speak in front of other people.
6. I worry my head will shake or nod in front of others.
APPENDIX D

THE STATE-TRAIT INVENTORY FOR COGNITIVE AND SOMATIC ANXIETY—STATE VERSION
Below is a list of statements which can be used to describe how people feel. Beside each statement are four numbers which indicate the degree with which each statement is self-descriptive of your mood at this moment (1 = not at all, 4 = very much so). Please read each statement carefully and write the number which best indicates how you feel right now, at this very moment, even if this is not how you usually feel.

1 = not at all
2 = a little
3 = moderately
4 = very much so

1. ____ My heart beats fast.
2. ____ My muscles are tense.
3. ____ I feel agonized over my problems.
4. ____ I think that others won’t approve of me.
5. ____ I feel like I’m missing out on things because I can’t make up my mind soon enough.
6. ____ I feel dizzy.
7. ____ My muscles feel weak.
8. ____ I feel trembly and shaky.
9. ____ I picture some future misfortune.
10. ____ I can’t get some thought out of my mind.
11. ____ I have trouble remembering things.
12. ____ My face feels hot.
13. ____ I think that the worst will happen.
14. ____ My arms and legs feel stiff.
15. ____ My throat feels dry.
16. ____ I keep busy to avoid uncomfortable thoughts.
17. ____ I cannot concentrate without irrelevant thoughts intruding.
18. ____ My breathing is fast and shallow.
19. ____ I worry that I cannot control my thoughts as well as I would like to.
20. ____ I have butterflies in the stomach.
21. ____ My palms feel clammy.
APPENDIX E

THE POSITIVE AND NEGATIVE AFFECT SCHEDULE—NEGATIVE AFFECT SCALE
This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you generally feel this way, that is, how you feel on average. Use the following scale to record your answers:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>very slightly or not at all</td>
<td>a little</td>
<td>moderately</td>
<td>quite a bit</td>
<td>extremely</td>
</tr>
</tbody>
</table>

___ distressed
___ upset
___ guilty
___ scared
___ hostile
___ irritable
___ ashamed
___ nervous
___ jittery
___ afraid
APPENDIX F

ORDERING AND ARRANGING BEHAVIORS INTERVIEW
1. What kinds of thoughts go through your mind while you are ordering and arranging your surroundings or when you are unable to order or arrange your surroundings?

2. Why do you think you engage in ordering and arranging behavior?

3. What do you think would happen if you were unable to order or arrange your surroundings?

4. Do you think that this behavior is excessive? How much of it do you think is acceptable? (Radomsky & Rachman, 2004a, p. 906).
APPENDIX G

VALIDITY QUESTIONS
The validity questions listed below used response options identical to the questionnaires in which they were embedded:

1. Please choose “agree” if you are paying attention right now.

2. I sometimes have fatal heart attacks while watching television.
APPENDIX H

PROCEDURE
Pre-select high/low ordering concerns individuals based on SCOPI-CR in mass screening

Re-administer SCOPI-CR to participants at start of session

Dismiss participants who have regressed to the mean

All participants: PANAS-NA, SIAS/SPS, OC-TCDQ, baseline SUDS

Randomization

Organized room: 3 minutes

Disorganized room: 3 minutes

All participants: Post-provocation SUDS
Then: Arranged items/urges to arrange items?

If "Yes": administer qualitative interview, debrief and dismiss

If "No": debrief and dismiss
APPENDIX I

DISORGANIZED ROOM STIMULUS
APPENDIX J

ORGANIZED ROOM STIMULUS
APPENDIX K

DATA LOSS IN CURRENT STUDY
Data collected from 211 individuals

78 individuals did not meet cut-offs (e.g., SCOPI-CR score between 16 and 25)
   n = 133

2 individuals failed at least one check question
   n = 131

4 individuals positive for OCD diagnosis or history
   n = 127

1 individual failed to complete questionnaires due to slow pace
Final N = 126