Obsessional beliefs and the implicit and explicit morality of intrusive thoughts

Bethany A. Teachman & Elise M. Clerkin

University of Virginia, Charlottesville, VA, USA

Version of record first published: 19 Jul 2007

To cite this article: Bethany A. Teachman & Elise M. Clerkin (2007): Obsessional beliefs and the implicit and explicit morality of intrusive thoughts, Cognition & Emotion, 21:5, 999-1024

To link to this article: http://dx.doi.org/10.1080/02699930600985576
Obsessional beliefs and the implicit and explicit morality of intrusive thoughts

Bethany A. Teachman and Elise M. Clerkin

University of Virginia, Charlottesville, VA, USA

Cognitive models of obsessive-compulsive disorder (OCD) suggest that misinterpreting intrusive thoughts exacerbates obsessional thinking. To evaluate this hypothesis, healthy participants \( (N = 91) \) were prompted to recollect their unwanted thoughts, and then beliefs about the immorality of these thoughts were manipulated. Next, participants completed implicit and explicit measures of self-evaluation and appraisals of unwanted thoughts. Results from structural regression analyses indicated that explicit responses to unwanted thoughts, such as evaluations of the significance of intrusive thoughts and state self-esteem, were predicted by pre-existing obsessional beliefs, but not by the morality instruction manipulation. In contrast, implicit responses, such as appraisals of unwanted thoughts as relatively important and evaluations of the self as relatively immoral and dangerous, were predicted by the interaction between specific obsessional beliefs (e.g., intolerance of uncertainty) and the morality instructions. Findings largely support cognitive models of OCD and suggest unique predictors of implicit and explicit responses to unwanted thoughts.

The highest stage in moral culture at which we can arrive is when we recognise that we ought to control our thoughts.

(Charles Darwin, author of *Origin of the Species*)

Despite the seemingly strange nature of obsessional thinking in obsessive-compulsive disorder (OCD), research suggests that the vast majority of the population has intrusive, unwanted thoughts that are similar in subject matter to the obsessions found in people with OCD (Rachman & de Silva, 1978; Salkovskis & Harrison, 1984). Based on this finding, cognitive models
of obsessions propose that it is not the content of obsessional thoughts, but the interpretation of the personal significance of the thoughts and their perceived consequences that cause the severe distress and repetitiveness characteristic of OCD (see excellent reviews in Frost & Steketee, 2002; Rachman, 1997; Salkovskis, 1996). According to cognitive models, OCD is thought to develop as a result of these maladaptive interpretations, which lead the vulnerable individual to interpret their normal intrusive thoughts as signs they are immoral, evil and dangerous (Rachman, 1997). There is mounting evidence to support these models—interpretations that suggest intrusive thoughts are important and personally significant, that one needs to be certain about the thoughts, and that imply one is responsible for harm differentiate persons with OCD from other anxiety groups and from non-anxious persons (Obsessive Compulsive Cognitions Working Group; OCCWG, 2001, 2003). Building on these important correlational findings, experimental investigations can now be conducted to evaluate the development of these interpretations and their consequences. In this way, we can more directly test the causal premise in these models—catastrophically misinterpreting unwanted thoughts exacerbates obsessional thinking and self-derision. The current study evaluates this hypothesis by manipulating beliefs about the immorality of unwanted thoughts and examining the impact on obsessional thinking and self-evaluation.

Recent studies with healthy (e.g., Rassin, Merckelbach, Muris, & Spaan, 1999; Teachman, Woody, & Magee, 2006), at-risk (e.g., Zucker, Craske, Barrios, & Holguin, 2002), and clinical obsessional samples (e.g., Forrester, Wilson, & Salkovskis, 2002) all suggest that OCD-relevant beliefs can be experimentally shifted with substantial consequences. For instance, increasing beliefs that one may be responsible for harm following intrusive thoughts has been tied to both increased anxiety and the urge to check (Ladouceur et al., 1995; Lopatka & Rachman, 1995; Shafran, 1997). Moreover, Forrester et al. (2002) found that simply providing participants with an intrusive thought about harm increased feelings of anxiety, blame, and the desire to take corrective action. Parallel investigations into manipulations of Thought–Action–Fusion (TAF), another OCD-relevant belief (see Rachman, Shafran, Mitchell, Trant, & Teachman, 1996; Shafran, Thordarson, & Rachman, 1996), suggest similar effects: Rassin et al. (1999) increased negative affect with a manipulation to exacerbate TAF beliefs, and Zucker et al. (2002) reduced anxiety with an intervention that diminished TAF beliefs. These findings provide strong support to cognitive models of OCD.

The current study builds on this exciting early work by manipulating beliefs about the morality of thoughts, and evaluating a wide range of potential consequences of holding these beliefs. Prior investigations have mostly relied on self-reported outcomes and focused predominantly on mood effects or results reflecting obsessional thinking and symptoms (e.g.,
urge to check or belief in responsibility for harm). Theoretically, there should also be consequences for self-evaluation, given that the major themes of obsessions (i.e., aggression, sex and blasphemy) are so central to moral systems (Rachman, 1997). As Rachman (1997) observes, obsessional patients interpret “these thoughts, impulses or images as revealing important but usually hidden elements in their character, such as: these obsessions mean that deep down I am an evil person, I am dangerous ... I am fundamentally immoral” (p. 794).

In addition to examining changes to self-evaluation following the manipulation of obsessional beliefs, we also consider responses to unwanted thoughts that are not self-reported by looking at both explicit and implicit attitudes. Explicit attitudes are those we can reflect upon and consciously endorse, whereas “implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects” (Greenwald & Banaji, 1995, p. 8). The rationale for examining both explicit and implicit responses in the current study follows from a number of sources. First, given the intense shame associated with obsessional thinking, it is not always easy for participants to report on these symptoms, so a measure that is less vulnerable to self-presentation concerns is helpful. Second, information-processing theories posit a key role for automatic processing of threatening information in anxiety (see McNally, 1995), particularly involuntary processing. Third, evidence from the social cognition field suggests that we may learn about different aspects of a person’s evaluation by looking at both implicit and explicit indicators. It is not that one is the “true” measure of evaluation, but at times implicit and explicit attitudes will agree and be highly related, while at other times they may diverge and be independent (Nosek, Banaji, & Greenwald, 2002). Importantly, they may relate to different anxiety-related concerns and triggers (e.g., Asendorpf, Banse, & Mücke, 2002), suggesting their unique predictive validity.

Supporting the utility of investigating both self-evaluation and implicit attitudes as consequences of obsessional thinking, Teachman et al. (2006) found divergent predictors of implicit and explicit outcomes following time spent thinking about one’s unwanted thoughts. Specifically, participants (screened to ensure they did not have high levels of OCD symptoms) were asked to identify an intrusive, unwanted thought they had previously experienced, and then were randomly assigned to receive information that unwanted thoughts were: (a) important and personally significant; (b) meaningless and not personally relevant; or (c) control instructions that did not speak to the importance of thoughts. After time spent reflecting on the unwanted thought and its significance, a series of responses were examined, including implicit and explicit measures of beliefs about the importance of unwanted thoughts and self-evaluation. Findings suggested that explicit
evaluations of unwanted thoughts and explicit self-esteem were predicted by
pre-existing obsessional beliefs (e.g., on the Obsessional Beliefs Question-
naire; OCCWG, 2001), but not by the instructions about the importance of
intrusive thoughts. In contrast, the implicit evaluation of unwanted thoughts
shifted in response to the instructions, and the implicit evaluation of the self
as dangerous was predicted by the interaction between pre-existing obses-
sional beliefs and the instructions. Further, the implicit and explicit measures
were not significantly correlated. These results suggest that implicit and
explicit responses to unwanted thoughts are reasonably independent and
have different predictors.

Results from Teachman et al. (2006) were interpreted to suggest that explicit responses might have been less vulnerable to the instruction
manipulation because they were under more strategic control. As a result,
explicit responses were more influenced by relatively stable, pre-existing
beliefs because the person had the opportunity to reflect on his or her
responses and tap into knowledge about his or her belief system (e.g., I am a
person who thinks unwanted thoughts are pretty meaningless). Thus, the
brief manipulation had little impact. In contrast, implicit responses are more
difficult to control. As a result, vulnerable individuals needed only a small
trigger to express more dysfunctional evaluations because the opportunities
for controlled correction were less available. We interpret the implicit
responses in line with models of cognitive vulnerability or diathesis-stress
models. For individuals who were vulnerable because of their pre-existing
obsessional beliefs, the instruction served as the spark that triggered these
beliefs and led to more negative self-evaluation. This finding is also
consistent with conclusions from a recent meta-analysis by Poehlman,
Uhlmann, Greenwald, and Banaji (2005), which suggested that implicit
measures may have higher predictive validity when considered in interaction
with self-reported attitudes.

Although this explanation helps explain the earlier results, it was a post
hoc explanation. Thus, replication of this pattern was needed. The current
study provides such an opportunity, and extends evaluation of the malleability of implicit and explicit responses to unwanted thoughts to a
new belief domain. Specifically, rather than shifting beliefs about the
importance of thoughts, the current study manipulates beliefs about the
morality of unwanted thoughts. To our knowledge, this is the first study to
experimentally investigate this belief (though see important correlational
work by Ferrier & Brewin, 2005). It is hypothesised that the manipulation
will shift reported beliefs about the morality of thoughts (this is in essence a
manipulation check), but change on other explicit responses is exploratory.
Instead, based on our earlier results (Teachman et al., 2006) and the strength
of pre-existing beliefs relative to the brief manipulation, we expect that pre-
existing beliefs will again be most influential in predicting explicit responses
to unwanted thoughts because these responses can be readily controlled. Implicit responses, on the other hand, are expected to be more vulnerable to the manipulation because responses on these measures are harder to strategically control. In line with our earlier findings and cognitive vulnerability models, this is hypothesised to be particularly true for individuals whose pre-existing belief system puts them at risk. Thus, an interaction is predicted whereby individuals with relatively high obsessional beliefs will have the most negative outcomes following the instructions about the immorality of unwanted thoughts.

It is predicted that a range of pre-existing obsessional beliefs can set the stage for maladaptive responses to unwanted thoughts, thus we will examine the sum of all obsessional beliefs (based on the Obsessional Beliefs Questionnaire) in interaction with the manipulation. It is also possible, however, that specific beliefs make one particularly vulnerable to different types of negative outcomes (i.e., shifts in self-evaluation versus judgements of unwanted thoughts). For example, it is plausible that believing unwanted thoughts are especially important will exacerbate judgements about the thoughts themselves, but have less influence on self-evaluation (because the belief really centres on the significance of the thought, rather than on the thinker). In contrast, beliefs that one should be able to control or needs to be certain about thoughts or actions might predict shifts in self-evaluation because these beliefs focus on the role of the thinker in having these unwanted thoughts. Thus, we will also consider how specific beliefs can set the stage for particular negative responses to unwanted thoughts.

**METHOD**

**Overall design**

All participants completed questionnaires and then used a list of commonly reported unwanted thoughts as a prompt to recollect some of their own unwanted thoughts. Participants who scored below the recommended clinical cutoff on the Obsessive Compulsive Inventory (OCI-R; Foa et al., 2002, see fuller description below) were invited to participate. They were randomly assigned to one of three conditions. In the Morally Significant condition, participants were informed that unwanted thoughts are meaningful and bear a strong connection to a person’s moral character. In the Meaningless condition, participants were informed that unwanted thoughts are meaningless and unrelated to a person’s moral character. Participants in the Control condition received no instructions about the meaning of unwanted thoughts. Following the instruction manipulation, participants rated their mood and completed implicit and explicit measures of self-evaluation and appraisals of their unwanted thoughts.
Participants

Undergraduate students and volunteers (N = 91; 51 women; mean age = 19.7 years, SD = 2.9) participated in the study, either for course credit as part of the psychology participant pool or for payment. Of these, 78% reported race or ethnicity as Caucasian, 8% African-American, 10% Asian, 1% Hispanic, and 1% biracial. Because of concerns that the manipulation might be harmful to participants who experienced clinical levels of OCD, only participants scoring below the recommended clinical cut-off on the OCI-R were invited to participate. The OCI-R was administered as part of a larger pre-screening battery for psychology students (who were eligible for the participant pool) at the outset of the semester.

Materials

**Pre-existing OCD and mood symptoms**

*Beck Depression Inventory (BDI).* The BDI (Beck & Steer, 1987) is a widely used self-report inventory of depression that has good reliability and validity. It is included in the current study to establish that the instruction condition groups do not differ on baseline depression levels. Cronbach’s alpha in the current sample was .84, suggesting good inter-item consistency.

*Obsessive-Compulsive Inventory – Revised (OCI-R).* The OCI-R (Foa et al., 2002) measures overall severity of OCD and is appropriate for assessing subclinical OCD symptoms in a nonclinical sample. The 18-item measure has good test–retest reliability, demonstrates convergent validity with other measures of OCD symptoms, and provides subscales for specific symptom domains (i.e., Washing, Checking, Ordering, Obsessing, Hoarding, and Neutralising). Foa et al. (2002) suggest that a cut-off score of 21 optimally distinguishes individuals at risk for OCD from those with no disorder. It is included in the current study to screen out individuals at risk for OCD and to establish that the instruction condition groups do not differ on baseline symptoms. Cronbach’s alpha from all participants who completed the screening measure was .90.

**OCD beliefs**

*Obsessional Beliefs Questionnaire-Short Form (OBQ).* The OBQ (OCCWG, 2005) is a 44-item questionnaire with good psychometric

---

1. Two participants were excluded from this final sample—one reported a prior diagnosis of OCD, and the other reported having no unwanted thoughts.
2. Only eight participants were not recruited through the psychology participant pool. There were no differences between the student and community groups on the key measures.
properties that assess general belief domains relevant to OCD (Cronbach’s alpha in the current study = .93). The total scale is used for most analyses, but follow-up tests will be conducted on subscales to evaluate specific obsessional beliefs. The short form is usually analysed based on three factors: (1) Responsibility/Threat Estimation; (2) Perfectionism/Certainty; and (3) Importance/Control of Thoughts. However, as the authors of the short form note, “retention of more subscales may be useful when investigators wish to examine the specific content of one or more of the hypothesised domains ... comparing it to related constructs” (OCCWG, 2005, p. 1538). Thus, given the interest in evaluating particular obsessional beliefs in the current study, the six separate belief subscales will be used (based on item assignment from OCCWG, 2001, 2003): Tolerance for Uncertainty, Threat Estimation, Control of Thoughts, Importance of Thoughts, Responsibility, and Perfectionism (Cronbach’s alpha in the current sample for subscales = .72, .78, .52, .84, .81, and .87, respectively).

**Responses to intrusive thoughts**

*Thought–Action–Fusion Scale (TAF).* The TAF (Shafran et al., 1996) is a 19-item self-report measure with adequate psychometric properties that assesses the tendency to equate thoughts and actions (e.g., believe that a negative thought is as harmful as the negative action). Only the 12-item moral TAF subscale (Cronbach’s alpha = .93) was used in the present study as a manipulation check to assess beliefs about the immorality of unwanted thoughts (e.g., “Having a blasphemous thought is almost as sinful to me as a blasphemous action”).

*State Self-Esteem Scale (SSES).* The SSES (Heatherton & Polivy, 1991) measures temporary changes in self-esteem and was used to determine the degree to which participants’ self-evaluation was related to their views on intrusive thoughts. The 20-item instrument has good reliability and validity (Cronbach’s alpha in the current sample was .93), and provides information about Performance, Social, and Appearance self-esteem.

*Personal Significance Scale (PSS).* The PSS (Rachman, 2001) is an 18-item instrument that measures beliefs about the personal meaning of one’s intrusive thoughts, such as whether unwanted thoughts imply that one is dangerous (e.g., “Do these thoughts mean you are a bad, wicked person?”). The PSS is scored by summing the 18 critical items mixed among five filler items, which are designed to make the intent of the scale less obvious. Cronbach’s alpha for the final scale was .90.
Positive and Negative Affect Scale – Expanded Form (PANAS-X). The PANAS-X (Watson & Clark, 1994) assesses current affect, and was included in the current study to check for mood effects following the manipulation. Only the 10-item General Negative Affect Scale was analysed in the present study (Cronbach’s alpha = .74).

Implicit Association Test (IAT). The IAT (Greenwald, McGhee, & Schwartz, 1998) measures automatic associations (in the sense that evaluations occur outside conscious control and, at times, outside awareness), thus appearing to share many of the qualities ascribed to anxiety schemata. The IAT has adequate psychometric properties, both in general (Greenwald & Nosek, 2001) and for the tasks used in the current study (see Teachman et al., 2006). Like many tasks used by social cognition researchers (Fazio, 2001), it is a reaction-time task that purportedly reflects strength of association between concepts. Specifically, the task involves comparing the time taken to classify stimuli when paired categories match a person’s automatic associations versus when paired categories contradict automatic associations. When categories are paired to match a person’s automatic associations, participants are expected to classify the stimuli more quickly. Thus, it is a relative measure of associations, rather than a measure of absolute evaluations.

In the present study, there were three IAT tasks. The first evaluated associations toward the self (versus others) paired with the concepts “Immoral” versus “Moral” (referred to as “IAT self-immoral”), the second evaluated associations between concepts of the self versus others with “Dangerous” versus “Harmless” (“IAT self-dangerous”), and the third evaluated associations between concepts of unwanted thoughts versus wanted thoughts with “Important” versus “Meaningless” (referred to as “IAT unwanted thoughts-important”). The IAT self-immoral task was always completed first, and then the other two IAT tasks were completed in random order following a brief break so that participants would not be slowed by fatigue. The IAT data were scored according to the scoring algorithm developed by Greenwald, Nosek, and Banaji (2003) to create a D score³ (no participants’ IAT data were deleted because of high error rates or unusually fast or slow response times). Overall IAT error rate was 5%. See Teachman et al. (2006) for a fuller description of the IAT tasks, as well as stimuli used within each category.

³ IAT D scores reflect the difference in mean reaction time across critical blocks divided by the standard deviations across blocks, which is conceptually similar to Cohen’s d (see Greenwald et al., 2003).
Procedure

The experimenter first provided the following definition of unwanted thoughts: “Unwanted thoughts are thoughts that pop into your head unexpectedly—thoughts you don’t really want to have. They are often socially unacceptable or contrary to how you try to live your life, and sometimes they can make people feel uncomfortable”. The experimenter indicated that the purpose of the study was to gain a better understanding of the meaning of those unwanted thoughts. Following informed consent, participants completed the OBQ and the BDI. Additionally, participants rated their state anxiety using a 0–100 verbal analogue scale, on which 100 reflects extreme anxiety. This served as a measure of baseline anxiety to ensure no differences between the groups prior to the manipulation. Next, participants read a lengthy list of intrusive thoughts that ordinary people had reported to researchers (e.g., harming someone I did not like, desire for sex with inappropriate partners). This list was adapted from Kyrios (2000) and Rachman and de Silva (1978). Participants were asked to read the list and reflect on thoughts that they had experienced; they were explicitly reassured that they would not be asked to verbally reveal which thought they had chosen.

Participants were then randomly assigned to the Morally Significant (N = 30), Meaningless (N = 30), or Control (N = 31) condition to receive different information about the significance of unwanted thoughts. Those who were assigned to the Morally Significant condition were informed:

Many people believe that unwanted thoughts like the ones on the list you read are an important window into the type of person you really are. The thoughts seem to be profoundly meaningful in either a literal or symbolic way. According to Thomas Jefferson, “No matter how a man may behave, it is what he thinks in the privacy of his mind that truly matters. We can learn much more about men from what they think rather than what they do”.

They were then given a mock survey containing a series of statements ostensibly conducted by the Psychology Department to study the morality of the student body (e.g., “I should be pure in my thoughts as well as my actions”). Participants were informed that fully 78% of those surveyed endorsed at least one of the survey statements, meaning that the student body at the university “does recognise that unwanted thoughts do bear a strong connection to their moral character”. Both the quote and the survey were fabricated for the purpose of shifting interpretations about the morality of intrusive thoughts.

Participants in the Meaningless condition were informed:

Many people believe that unwanted thoughts like the ones on the list you read are meaningless fluctuations of the mind, unrelated to the type of person you really are.
The thoughts are not meaningful in either a literal or a symbolic way. We do not think these thoughts reflect your personal values or your moral character. According to Thomas Jefferson, “No matter what a man may think in the privacy of his mind, it is how he behaves that truly matters. We can learn much more about men from what they do rather than what they think”.

They were given the same mock survey as those in the Morally Significant condition, but were informed that only 22% endorsed at least one of the survey statements, meaning that the student body “does recognise that unwanted thoughts do not bear a strong connection to their moral character”.

Participants assigned to the Control condition were not given any information about the meaning of unwanted intrusive thoughts. However, all participants were informed: “We have a computer task that we are developing to measure aspects of cognition that could be important links to the meaning of these thoughts. To get ready for the task, I would like you to spend the next two minutes thinking carefully about the unwanted intrusive thoughts that have occurred most frequently for you or that have been most upsetting”. Additionally, those in the Morally Significant and Meaningless conditions were told to think about these unwanted thoughts in light of whether they meant anything about a person’s virtue. In this way, the conditions were matched for time spent thinking about their unwanted thought.

After the two-minute thought period, participants completed the PANAS, TAF-moral, and PSS, followed by the IAT self-immoral task. Next, there was a brief reminder of the condition-specific instructions to reinstate the manipulation, and the SSES was administered. Participants then received a final reminder of their instruction condition before completing the remaining two IAT tasks (IAT self-dangerous and IAT unwanted thoughts-important, in counterbalanced order). Lastly, participants were fully debriefed and assured that intrusive thoughts are a normal phenomenon.

RESULTS

Descriptive statistics

Means and standard deviations for the measures of depression, OCD symptoms and beliefs are listed in Table 1. Analysis of variance (ANOVA) tests showed no significant group differences on the BDI, $F(2, 88) = 0.89, p > .10, f = 0.14$, OCI-R, $F(2, 87) = 0.004, p > .10, f = 0.009$, or OBQ, $F(2, 88) = 0.99, p > .10, f = 0.15$, suggesting equivalence across instruction groups following random assignment. Further, correlations among the mood and OCD measures were all significant in the expected, positive direction (with
the exception of the correlation between the BDI and OCI-R, which did not reach significance; see Table 2). Finally, chi-square tests indicated the groups did not differ by gender ($\chi^2 = 3.18, p > .10$) or ethnicity ($\chi^2 = 15.80, p > .10$).

### Manipulation check

To evaluate whether the instruction manipulation was effective at changing beliefs about the morality of intrusive, unwanted thoughts, the TAF-Moral measure was compared across groups. The Morally Significant group was expected to report a greater morality bias following the instruction manipulation than either the Control or Meaningless groups. As hypothesized, the planned weighted contrast (Morally Significant: $+2$, Control: $-1$, Meaningless: $-1$) showed significant differences, $t(88) = 2.27$, $p = .03$, $d = 0.48$), indicating that the manipulation was successful. Following Teachman et al. (2006), these weightings were based on the hypothesis that individuals with few OCD symptoms would evaluate their unwanted thoughts as being fairly meaningless in the absence of other information (so the Control and Meaningless conditions were expected to be similar to each other, but different from the Morally Significant condition).

### Table 1

**Descriptive statistics for mood and OCD symptoms**

<table>
<thead>
<tr>
<th>Measure</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beck Depression Inventory</td>
<td>5.92</td>
<td>4.96</td>
</tr>
<tr>
<td>Obsessive-Compulsive Inventory—Revised</td>
<td>10.24</td>
<td>4.77</td>
</tr>
<tr>
<td>Obsessional Beliefs Questionnaire—Short Form</td>
<td>142.32</td>
<td>32.81</td>
</tr>
<tr>
<td>Interpretation of Intrusions Inventory</td>
<td>81.30</td>
<td>49.38</td>
</tr>
</tbody>
</table>

### Table 2

**Correlations among mood, OCD symptoms, and IAT measures**

<table>
<thead>
<tr>
<th></th>
<th>$OCI$-$R$</th>
<th>$OBQ$</th>
<th>$BDI$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obsessive-Compulsive Inventory—Revised (OCI-R)</td>
<td></td>
<td>.21*</td>
<td>.14</td>
</tr>
<tr>
<td>Obsessional Beliefs Questionnaire—Short Form (OBQ)</td>
<td></td>
<td></td>
<td>.56**</td>
</tr>
<tr>
<td>IAT unwanted thoughts-important</td>
<td>-.03</td>
<td>.17</td>
<td>.21*</td>
</tr>
<tr>
<td>IAT self-immoral</td>
<td>-.04</td>
<td>-.01</td>
<td>.01</td>
</tr>
<tr>
<td>IAT self-dangerous</td>
<td>-.01</td>
<td>-.18*</td>
<td>-.07</td>
</tr>
</tbody>
</table>

*Notes: **$p < .01$ level; *$p < .05$ level; +$p < .10$. BDI = Beck Depression Inventory. IAT refers to Implicit Association Test.*
Obsessional beliefs and responses to unwanted thoughts

It was expected that pre-existing obsessional beliefs would be associated with more negative responses following time spent thinking about unwanted thoughts. Results supported this hypothesis, with positive correlations between obsessional beliefs and evaluations of the personal significance (PSS with OBQ: $r = .57, p < .001$) and morality of unwanted thoughts (TAF-moral with OBQ: $r = .53, p < .001$), as well as more negative affect (PANAS with OBQ: $r = .49, p < .001$). Further, obsessional beliefs were related negatively to state self-esteem (SSES with OBQ: $r = -.59, p < .001$; negative relationships were expected because higher scores on the SSES indicate more positive self-esteem). In contrast, the OBQ was not significantly related to any of the IAT tasks (see Table 2; all $r < .18, p > .10$, with the exception of one non-significant trend between the OBQ and IAT self-dangerous: $r = -.18, p = .08$). These results replicate Teachman et al. (2006), suggesting robust relationships between pre-existing obsessional beliefs and explicit responses to unwanted thoughts, but little relationship with implicit responses.

Impact of instruction condition

Recall that the instructions were mainly expected to shift implicit responses in interaction with pre-existing beliefs. However, main effects were also assessed. To explore whether the Morally Significant group, relative to the other groups, would experience a more negative reaction following time spent thinking about unwanted thoughts, planned weighted contrasts (Morally Significant: +2, Control: −1, Meaningless: −1) were examined for each of the implicit and explicit measures of self-evaluation and responses to unwanted thoughts (see Table 3 for means and standard deviations). The planned contrasts were all non-significant, SSES: $t(86) = 1.04, p > .10, d = 0.22$, PSS: $t(88) = 0.50, p > .10, d = 0.11$, IAT unwanted thoughts-important: $t(87) = 0.45, p > .10, d = 0.10$, IAT self-immoral: $t(87) = 0.91, p > .10, d = 0.20$, IAT self-dangerous: $t(87) = 0.63, p > .10, d = 0.14$. In addition, the equivalent planned contrast for the PANAS negative affect subscale indicated no significant mood effect, $t(88) = 1.50, p = .14, d = 0.32$. These findings suggest that information about the morality of unwanted thoughts shifted the explicit moral evaluation (see TAF-Moral results above), but not mood or the other implicit or explicit responses.4

---

4 These findings replicate those found by Teachman et al. (2006) regarding little effect of the manipulation as a main effect, with the exception of the results for the IAT unwanted thoughts-important task.
Interaction between pre-existing obsessional beliefs and instruction condition

Modelling plan. It was expected that there would be an interaction between pre-existing beliefs and instruction condition on implicit responses. To evaluate this question, a structural regression model was first examined for the full sample (see Figure 1a), which allows for simultaneous consideration of relationships between the predictor (obsessional beliefs, as measured by the OBQ) and dependent variables (the three implicit measures; see discussion of structural equation modelling in McArdle & Hamagami, 1996). A decision was made to predict each implicit measure independently, rather than creating a latent implicit factor because relationships among the implicit indicators were variable (IAT self-immoral with IAT unwanted thoughts-important, $r = .71, p < .05$; IAT self-immoral with IAT self-dangerous, $r = .28, p < .02$; IAT unwanted thoughts-important with IAT self-dangerous, $r = .34, p > .10$), and because the implicit responses were thought to share method variance but still reflect different constructs. All models were fit to the data using AMOS, and full information maximum likelihood methods were used so that incomplete data were treated as missing at random (Little & Rubin, 1987).

The analysis with the full sample allowed for a test of the main effect of pre-existing obsessional beliefs predicting implicit responses. This main effect was not expected to be significant (based on Teachman et al., 2006, and the
correlational analyses noted above), but was conducted as a first step to set up the interaction with instruction condition, the primary question of interest. To evaluate the interaction, multi-group analyses were carried out to examine differences across conditions for the model. This test allows us to examine interactions between obsessional beliefs and instruction condition for each implicit measure simultaneously. In the multi-group analyses, the change in model fit was evaluated relative to a baseline model where all regressions were estimated freely (unconstrained) across the groups. In the constrained model, the structural weights (regression paths from the OBQ to each implicit response) were constrained equal across the instruction groups. Where a significant change in fit was observed, follow-up analyses were then conducted to evaluate the source of the interaction by sequentially releasing one

**Figure 1.** Structural regression models for full sample with implicit (Figure 1a) and explicit (Figure 1b) responses to unwanted thoughts predicted by obsessional beliefs (on the OBQ). *Note:* IAT refers to Implicit Association Test. Standardised parameter estimates that are significant at $p < .05$ are noted with solid lines.
constrained parameter at a time to determine whether it had been the cause of
the change in fit (i.e., looking to see if the loss of fit was no longer significant
when groups were not constrained equal).

Next, parallel analyses were run to evaluate the main effect of pre-existing
obsessional beliefs and the interaction between beliefs and instruction
condition on explicit responses. The opposite pattern of results was predicted
for these analyses, relative to the implicit response models; namely, a main
effect for prediction by pre-existing beliefs but no interaction with instruction
condition was expected. To match these analyses to the models tested for
implicit responses, the explicit measures were predicted independently, rather
than creating a latent explicit factor, though relationships among the explicit
indicators were moderate (TAF-Moral with PSS, \( r = .53, p < .001 \); TAF-
Moral with SESS, \( r = -.40, p < .001 \); SSES with PSS, \( r = -.52, p < .001 \)).

For the models testing main effects, assessment of model fit was based on
the root-mean-square error of approximation index (RMSEA; values less
than 0.08 can be considered an acceptable fit as defined by Browne &
Cudeck, 1993), and the normed fit index (NFI; values above 0.90 indicate an
acceptable fit; Hu & Bentler, 1999). For the multi-group analyses, the
primary outcome was the change in fit (\( \Delta \chi^2 \) on relevant \( \Delta df \) distribution)
between the baseline and constrained models.

Evidence for model fit. Table 4 outlines the goodness-of-fit indices for
each model, and the change in fit between models. For the models testing
main effects of the OBQ predicting responses to unwanted thoughts, neither
the implicit nor explicit model goodness-of-fit indices suggest a strong
overall fit. However, as expected, all paths from the OBQ to the explicit
responses were significant (and the NFI suggested a reasonable fit for this
model) and none of the paths to the implicit responses were significant at
\( p < .05 \). Figure 1 outlines the standardised regression weights for both
models for the full sample. These results support the hypothesis of a main
effect for pre-existing obsessional beliefs predicting explicit but not implicit
responses to unwanted thoughts.

Results for the interaction effects suggest a more complicated story. As
predicted, there was no evidence for an interaction effect between the OBQ
total score (or any OBQ subscales) and instruction condition predicting
explicit responses to unwanted thoughts. Table 4 notes the non-significant
change in fit (see \( \Delta \chi^2 \) on \( \Delta df \) ) between the baseline and constrained model.
In contrast to hypotheses, the model testing the interaction between the
OBQ total and implicit responses was also not significantly different from
the baseline, unconstrained model. However, follow-up exploratory analyses
examining the OBQ subscales separately suggested that each of the implicit
responses was predicted by an interaction between an OBQ subscale and
instruction condition.
TABLE 4
Fit statistics for structural regression models

<table>
<thead>
<tr>
<th>Goodness-of-fit indices</th>
<th>( \chi^2 )</th>
<th>( df )</th>
<th>( A\chi^2 ) on ( Adf )</th>
<th>ANFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect (full sample)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit responses predicted by obsessional beliefs (Figure 1a)</td>
<td>6.88</td>
<td>3</td>
<td>—</td>
<td>.45</td>
<td>.120</td>
</tr>
<tr>
<td>Explicit responses predicted by obsessional beliefs (Figure 1b)</td>
<td>17.46</td>
<td>3</td>
<td>—</td>
<td>.86</td>
<td>.231</td>
</tr>
<tr>
<td><strong>Interaction Effect (OBQ-Total x Instruction Condition)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit responses—baseline (unconstrained)</td>
<td>15.75</td>
<td>9</td>
<td>Baseline</td>
<td>.36</td>
<td>.092</td>
</tr>
<tr>
<td>Implicit responses—structural weights constrained</td>
<td>19.80</td>
<td>15</td>
<td>4.05 on 6</td>
<td>.19</td>
<td>.060</td>
</tr>
<tr>
<td>Explicit responses—baseline (unconstrained)</td>
<td>24.40</td>
<td>9</td>
<td>Baseline</td>
<td>.82</td>
<td>.139</td>
</tr>
<tr>
<td>Explicit responses—structural weights constrained</td>
<td>26.88</td>
<td>15</td>
<td>2.48 on 6</td>
<td>.80</td>
<td>.095</td>
</tr>
<tr>
<td><strong>Interaction Effect (OBQ-Tolerance of Uncertainty x Instruction Condition)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit responses—baseline (unconstrained)</td>
<td>13.88</td>
<td>9</td>
<td>Baseline</td>
<td>.52</td>
<td>.078</td>
</tr>
<tr>
<td>Implicit responses—structural weights constrained</td>
<td>26.43</td>
<td>15</td>
<td>12.55 on 6*</td>
<td>.09</td>
<td>.093</td>
</tr>
<tr>
<td>IAT self-dangerous free (other weights constrained)</td>
<td>19.56</td>
<td>13</td>
<td>5.69 on 4</td>
<td>.32</td>
<td>.076</td>
</tr>
<tr>
<td>IAT self-immoral free (other weights constrained)</td>
<td>22.04</td>
<td>13</td>
<td>8.16 on 4</td>
<td>.24</td>
<td>.089</td>
</tr>
<tr>
<td><strong>Interaction Effect (OBQ-Importance of Thoughts x Instruction Condition)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit responses—baseline (unconstrained)</td>
<td>14.04</td>
<td>9</td>
<td>Baseline</td>
<td>.58</td>
<td>.080</td>
</tr>
<tr>
<td>Implicit responses—structural weights constrained</td>
<td>27.59</td>
<td>15</td>
<td>13.55 on 6*</td>
<td>.17</td>
<td>.098</td>
</tr>
<tr>
<td>IAT unwanted thoughts-important free (other weights constrained)</td>
<td>18.54</td>
<td>13</td>
<td>4.51 on 4</td>
<td>.44</td>
<td>.070</td>
</tr>
</tbody>
</table>

Notes: *Change in model fit relative to baseline is significant at \( p \leq .05 \) on \( \chi^2 \) distribution with relevant \( Adf \). NFI = Normed Fit Index; RMSEA = root-mean-square error of approximation.
For these analyses, the identical multi-group model comparison outlined for the OBQ total was re-run for each subscale to determine whether particular obsessional beliefs interacted with condition. Again, the outcome of interest was whether the change in fit between the baseline model and the model with structural weights constrained across groups was significant. As noted in Table 4, both the OBQ Tolerance of Uncertainty and Importance of Thoughts subscales interacted with instruction condition, such that constraining the structural weights equal across groups resulted in a significant loss of fit relative to the baseline model where all paths were estimated freely. For these OBQ subscale models, follow-up analyses were then performed to determine the source of the interaction by freeing each implicit measure while constraining the other two to evaluate when the loss in fit was no longer significant compared to baseline. Paths that were the source of the significant interactions are noted in Table 4.

Results for the OBQ Tolerance of Uncertainty analysis suggested interactions with instruction condition for both the IAT self-dangerous and self-immoral tasks. For the IAT self-dangerous task, subsequent analyses freeing one instruction condition at a time (and leaving the other two constrained equal) indicated that the Morally Significant and Meaningless groups differed significantly from one another (with the Control group intermediate between the two and not significantly different from either group). This interaction is displayed in Figure 2. For the Meaningless instruction group, there is a significant negative slope, such that if you hear that your unwanted thoughts are meaningless, then you evaluate yourself as less implicitly dangerous (relative to harmless) the more you endorse beliefs about being intolerant of uncertainty (standardised regression weight = -.44; \( p = .008 \)). The regression weight for the Morally Significant group was in the opposite direction, but did not reach significance (standardised regression weight = .19; \( p > .10 \)). This intriguing interaction may occur because being told unwanted thoughts are meaningless resolves uncertainty in a good way, so people who are less tolerant of uncertainty may then feel better about themselves. Supporting this idea, the significant interaction for the IAT self-immoral task followed a very similar pattern (see Figure 2). The meaningless group again had a negative slope (standardised regression weight = -.32; \( p = .07 \)), and the slopes for the Morally Significant and Control groups were not different from one another. These subscale analyses were exploratory so should be interpreted with caution, but the pattern for the meaningless group was consistent for both indicators of implicit self-evaluation.

Examining the interaction for the OBQ Importance of Thoughts subscale revealed that the IAT unwanted thoughts-important task was the critical indicator (as Table 4 indicates, the change in fit from the baseline model is no longer significant after freeing that path across instruction conditions). Follow-up analyses freeing one instruction condition at a time suggested that
Figure 2 (above and opposite). Interaction effects: (a) IAT self-dangerous × OBQ Tolerance of Uncertainty scores; (b) IAT self-immoral × OBQ Tolerance of Uncertainty scores; and (c) IAT unwanted thoughts-important × OBQ Importance of Thoughts scores as a function of instruction condition. Note: Higher scores on the OBQ Tolerance for Uncertainty subscale reflect greater intolerance.
all three instruction groups were significantly different from one another. The Morally Significant group had a significantly positive slope (standardised regression weight \( .44; p = .009 \)), suggesting that implicit and explicit beliefs about the importance of unwanted thoughts are positively related. The pattern was similar for the Meaningless group (though it did not reach standard levels of significance; standardised regression weight \( .31; p = .08 \)), but the control group had a non-significant negative slope (standardised regression weight \( -.33; p = .06 \); see Figure 2). Recall that the control group was the only condition not cued to think about the morality of their thoughts. These results suggest that, regardless of the specific instructions, being primed to think about the moral significance of unwanted thoughts strengthens the implicit/explicit relationship.

**DISCUSSION**

Cognitive models of OCD suggest that it is the catastrophic misinterpretation of intrusive thoughts that exacerbates obsessional thinking and leads to negative self-evaluation. To evaluate this hypothesis, beliefs about the immorality of unwanted thoughts were manipulated in a sample selected for low OCD symptoms. Results indicated that the manipulation was
successful, as denoted by higher TAF-moral scores for the group instructed that unwanted thoughts reflect one’s moral character. In addition, structural regression analyses indicated that explicit responses to unwanted thoughts, such as evaluations of the significance of intrusive thoughts and state self-esteem, were predicted by pre-existing obsessional beliefs, but not by the instruction condition. In contrast, implicit responses, such as appraisals of unwanted thoughts as important and evaluations of the self as relatively immoral and dangerous, were predicted by the interaction between specific obsessional beliefs (though not by overall severity of one’s obsessional beliefs) and the instructions provided about the morality of unwanted thoughts. Replicating earlier results from Teachman et al. (2006), these findings suggest that obsessional beliefs can be shifted in non-obsessional participants and that there are distinct predictors of implicit and explicit responses to unwanted thoughts.

These findings offer considerable support for cognitive models of OCD. The observation that pre-existing beliefs (that are thought to be vulnerability factors for OCD symptoms) predicted more disapproving self-evaluation, harsher judgements of the immorality and significance of unwanted thoughts, and greater negative affect following time spent thinking about intrusive thoughts is in line with theoretical predictions. It is especially interesting that these beliefs were predictors even in a healthy sample. Longitudinal studies will be important to determine the potential role of these beliefs in the development of more serious obsessional symptoms.

Although experiencing an unwanted thought is a normal event, receiving direct information about the moral significance of that thought is not typical. Perhaps this may partly explain how such a brief manipulation with a healthy sample can have substantial consequences. The manipulation check, the moral subscale of the Thought–Action–Fusion scale, supported the idea that the instructions successfully shifted beliefs about the immorality of having intrusive thoughts and their equivalence to immoral actions. It is possible, however, that this shift reflects demand effects, so it is notable that this manipulation also interacted with pre-existing obsessional

5 Results across Teachman et al. (2006) and the current study showed very similar patterns. In both cases, pre-existing obsessional beliefs predicted all explicit (but not implicit) responses to unwanted thoughts, whereas the interaction between pre-existing obsessional beliefs and the instruction conditions predicted implicit (but not explicit) responses. Notwithstanding, there were some differences, including the finding of a main effect of instruction condition on the IAT unwanted thoughts-important task in Teachman et al. (2006), but not in the current study. In addition, the OBQ total score interacted with instructions condition to shift the IAT self-dangerous task in the earlier study (unfortunately, interactions with specific belief domains were not examined in that study), whereas it was specific belief domains (and not the total) that were predictive in the current study. Also, there were significant interactions observed for each of the IAT tasks in the present study (rather than with only one IAT task).
beliefs to shift *implicit* (less controllable) self-evaluation. In particular, participants who started the study with greater beliefs related to intolerance of uncertainty evaluated themselves as relatively less immoral and dangerous when they were assured their unwanted thoughts were meaningless with respect to their moral character. It is worth noting that the manipulation changed self-evaluation for the meaningless group, but not for the group instructed that their thoughts were morally significant (the slope was in the opposite direction for this group but did not reach significance). This may reflect the relatively healthy nature of this sample, such that they are more readily convinced that unwanted thoughts are *not* meaningful than that they are sinful (though see Rassin et al., 1999).

It is important to point out that most responses did not shift following the moral significance instructions—it was only in interaction with specific, pre-existing obsessional beliefs (namely, the need to be certain and the importance of thoughts, rather than the other four belief domains) that the implicit responses were different across instruction groups. As outlined in the introduction, examining the main effect of instruction condition was an exploratory aspect of the study given the limited findings from Teachman et al. (2006) and the clearer rationale for expecting the instructions to be predictive in interaction with pre-existing obsessional beliefs. We suspect that the brevity of the manipulation, relative to the individual’s pre-established moral system, would make it difficult to shift responses that are relatively easy to control strategically. Moral views are likely well-elaborated and central to the self, meaning that they may be particularly resistant to persuasion (Zuwerink & Devine, 1996). Instead, it is those responses that occur outside conscious control that will be most responsive to the manipulation given the appropriate susceptibility. This explanation fits well with models of cognitive vulnerability (e.g., Clark & Beck, 1999) and diathesis-stress or mood-state-dependent models (e.g., Persons & Miranda, 1992). In this sense, beliefs related to the need to be certain and to the importance of thoughts set the stage for activation of the implicit responses following the instruction trigger. It will be important in future work to figure out why the other obsessional belief domains—Threat Estimation, Control of Thoughts, Responsibility, and Perfectionism—did not result in significant interactions with the instruction manipulation. Our suspicion is that distinctions among specific obsessional beliefs are critical because they may differentially predict shifts in self-evaluation versus shifts in beliefs about the thought itself (e.g., that it is important or will come true, etc.), and the different belief domains may also be more or less problematic for a given type of intrusive thought. This highlights the likely value of targeting the
range of problematic obsessional beliefs in cognitive therapy for OCD (Wilhelm & Steketee, 2006).

We did not predict that beliefs related to intolerance of uncertainty would be unique predictors a priori, so we can only speculate at this stage about why they were especially influential in shifting implicit self-evaluation. There is little question that uncertainty is central to OCD phenomenology, and difficulties accepting uncertainty and doubt have long been recognised as hallmarks of OCD (Janet, 1908; cited in Ey, Bernard, & Brisset, 1963). It seems likely to us that the tie between morality and uncertainty may be particularly strong because there is no definitive answer for many moral questions, so they are ripe to trigger distress among persons who are uncomfortable with ambiguity. Evaluating the moral nature of thoughts is a particularly challenging domain for people intolerant of uncertainty because although we rarely receive clear instructions about the moral significance of our unwanted thoughts, we are nonetheless frequently confronted with mixed messages relevant to this topic. For instance, religious texts vary considerably in their evaluation of the sinfulness of thoughts (e.g., “But I tell you that anyone who looks at a woman lustfully has already committed adultery with her in his heart”; Matthew, 5: 28, Holy Bible, New International version). This may partly explain the positive relationship observed between intense religiosity and higher levels of OCD symptoms and obsessional beliefs (e.g., Rassin & Koster, 2003), and the finding of greater religion-oriented obsessions in some religions (e.g., Protestants and Catholics) relative to others (e.g., Jews; see Abramowitz, Huppert, Cohen, Tolin, & Cahill, 2002). In contrast to certain religious teachings about the immorality of thoughts, Thomas Jefferson suggested, “The only exact testimony of a man is his actions, leaving the reader to pronounce on them his own judgment” (Jefferson to L. H. Girardin, 1815). Thus, the morality of intrusive thoughts is likely uncertain to many people, making it an uncomfortable question for those who desire immediate resolution to ambiguity.

The other significant interaction between pre-existing beliefs and instruction condition reflected convergence between implicit and explicit evaluations of the importance of unwanted thoughts. Specifically, participants who were primed about the moral import of their unwanted thoughts indicated higher implicit evaluation of the relative importance of their unwanted thoughts the more they also held these beliefs explicitly. Interestingly, the pattern was the same regardless of whether participants were instructed that unwanted thoughts were significant or meaningless (though the Morally Significant group had a significantly more positive slope than the Meaningless condition). This basic similarity across conditions was an unexpected finding in that both the Morally Significant and
Meaningless groups showed the same pattern despite opposite instructions. However, this pattern fits well with models that suggest greater convergence between implicit and explicit evaluations as a function of elaboration (e.g., Nosek, 2005). It appears that simply priming beliefs related to the meaning of unwanted thoughts, regardless of the specific nature of the prime, can activate implicit evaluations that are in line with explicit, consciously endorsed beliefs.

Limitations and conclusion

One limitation of the current study was that immediate behavioural intentions were not investigated. In a non-clinical sample, one would not expect substantial rates of overt compulsive behaviours, but efforts to suppress the thought and strong urges to neutralise would be anticipated among participants who believe their unwanted thoughts are immoral. It will also be helpful in future work to include other implicit measures of responses to unwanted thoughts, given that the IAT is constrained in the sense of being a relative measure, and there is some debate about the processes underlying IAT effects (e.g., Rothermund & Wentura, 2004). Finally, perhaps the most significant limitations of the current study were the brevity of the manipulation and the inclusion of only those participants who were low in OCD symptoms. We felt these restrictions were necessary for ethical reasons (until we understand the impact of exacerbating obsessional beliefs). However, this makes it difficult to generalise from the current findings to the development of obsessional thinking over time or to clinical populations, obviously important questions given cognitive models of OCD.

Notwithstanding, the current findings do shed light on the impact of obsessional beliefs, such as intolerance of uncertainty and judgements of the moral significance of unwanted thoughts, and how these beliefs contribute to maladaptive appraisals and negative self-evaluation following exposure to intrusive thoughts. These results replicate earlier findings from Teachman et al. (2006) regarding the different predictors of implicit and explicit responses to unwanted thoughts, and point to the value of investigating automatic cognitive processing that lies outside conscious control. OCD is a complicated disorder and the beliefs can seem magical or even delusional at times, but the current results support cognitive models that highlight the normative nature of intrusive thoughts. Even in healthy samples, small manipulations about the significance of unwanted thoughts can have serious consequences. More directly investigating the clinical implications of these beliefs for onset and reduction of OCD symptoms will be an exciting next step.
REFERENCES


