Supplementary Information

Supplementary Methods

Drift diffusion model. In order to aid sampling and reduce correlation in the sampled posterior, we divided $drift_c$ into a baseline parameter, $drift_{c=0H}$, representing the 0H condition, and the remaining parameters, $drift_{c'}$, which were fit relative to this baseline. The same was done for $drift_{c,padua}$ and the corresponding two boundary parameter sets. Parameters reported in the paper are the original parameters for each individual condition (e.g. the reported parameter for $drift_{c=0V} = drift_{c=0H} + drift_{c'=0V}$). Broad weakly informative priors were used for all parameters; this was $N(0, 20)$ for most parameters, as described below. The prior for each of $drift_{c=0H}$, $drift_{c'}$, $drift_{c=0H,padua}$, $drift_{c',padua}$, $\sigma_{drift,trial}$, $bound_{c'}$, $bound_{c=0H,padua}$, $bound_{c',padua}$, $bias_{\mu}$ (the single bias mean), $bias_{padua}$, and $\sigma_{bias,trial}$ was $N(0, 20)$. $\sigma_{drift,trial}$ and $\sigma_{bias,trial}$ also had a lower bound of 0. The prior for $bound_{c=0H}$ was $N(1, 20)$, biased positive because boundary separation must be positive, but with a large variance, and was constrained to be greater than or equal to 0. The priors for $drift_s$, $bound_s$, and $bias_s$ were hierarchical, $N(0, \sigma_{drift})$, $N(0, \sigma_{bound})$, and $N(0, \sigma_{bias})$, respectively, with a prior of $N(0, 20)$ for each σ. The prior for each subject's non-retrieval/decision time was also hierarchical, $N(\tau_{\mu}, \tau_{\sigma})$, with the prior for τ_{μ} set to $N(0.25, 5)$, which similar to the prior for the baseline boundary separation was biased positive but with a large variance, and the prior for τ_{σ} was set to $N(0, 5)$. All τ parameters were constrained to be greater than or equal to 0.

Metacognitive model. Similar to $drift_c$ and $bound_c$, in order to reduce correlation in the sampled posterior and aid sampling, $\beta_{c=0H,padua}$ was defined to represent condition $0H$, and the remaining sampled parameters, $\beta_{c',padua}$, were added to this baseline. Reported parameters are the original parameters (e.g. $\beta_{c=0V,padua} = \beta_{c=0H,padua} + \beta_{c'=0V,padua}$). $\beta_{c,performance}$, $\beta_{c=0H,padua}$, $\beta_{c',padua}$, $\beta_{c,correct}$, and $\beta_{c,padua,correct}$ had $N(0, 20)$ priors. Because $\beta_{c,performance,correct}$ and $\beta_{c,performance}$ were anti-correlated in the posterior, in order to aid sampling we divided the former as $\beta_{c,performance,correct} = -\beta_{c,performance} + \beta_{c',performance,correct}$, setting the prior for each of $\beta_{c',performance,correct}$ to $N(0, 20)$. As for the other re-parameterizations, we report values for the original $\beta_{c,performance,correct}$. The prior for $\beta_{s,correct}$ was hierarchical, $N(0, \sigma_{\beta_{s,correct}})$, with the prior for $\sigma_{\beta_{s,correct}}$ set to $N(0, 20)$.

We also re-parameterized each of $cutpoint^1_c$ and $cutpoint^2_c$ relative to a baseline which represented condition $0H$. The priors for $cutpoint^1_{c=0H}$, $cutpoint^1_{c'}$, $cutpoint^2_{c=0H}$, and $cutpoint^2_{c'}$ were $N(0, 20)$. The priors for $cutpoint^1_s$ and $cutpoint^2_s$ were hierarchical, $N(0, \sigma_{cutpoint^1})$ and $N(0, \sigma_{cutpoint^2})$, respectively, with the priors for $\sigma_{cutpoint^1}$ and $\sigma_{cutpoint^2}$ set to $N(0, 20)$.

Supplementary Results

Padua subscales. The Padua Inventory has five subscales: checking compulsions, contamination obsessions and washing compulsions, dressing and grooming compulsions, obsessional thoughts of harm to self and others, and obsessional impulses to harm self and others. It is of obvious interest whether the results differ by symptom cluster, and in particular whether the results are specific to checking symptoms. In practice, however, the subscales are highly

correlated, and asking such a question poses both methodological and conceptual challenges. Tables S2 and S3 display the correlation matrices between subscales for the two datasets. As an exploratory analysis, we tested how discriminability, confidence bias, and confidence sensitivity related to scores on each individual measure. All results were similar to using the aggregate scores across all subscales, as reported in the main text. It would be of interest to attempt to test the specificity of the effects using another instrument which could potentially better distinguish symptom clusters (e.g. Abramowitz & Deacon, 2006). However, it should be noted that it is not fully understood the extent to which groups of symptoms and the underlying cognitive and neural mechanisms driving them are truly independent. For example, contamination obsessions and washing compulsions may be special cases of checking, rather than completely independent symptoms. Thus there is not a completely clear a priori reason to believe the results would be specific to checking, even using an instrument that can better distinguish symptoms. Instead, whatever the results, they would add to this debate.

Bias in memory strength criterion. Another information processing dimension of interest is the drift criterion (Ratcliff & McKoon, 2008; C. N. White & Poldrack, 2014): the reference relative to which drift rates are considered positive and directed toward the *old* boundary, or negative and directed toward the *new* boundary. More specifically, it is assumed that:

$$\text{total drift rate} = \text{signal strength} - \text{criterion}.$$

A smaller criterion amounts to a lower memory strength requirement for judging an event to be old. While we are not aware of a fully systematic dataset on the prevalence of OCD symptoms, at least anecdotal clinical description suggests symptoms that would be consistent with either a lower or higher drift criterion (Wells, 2002). For example, individuals may be concerned with both events such as whether a door was locked (which might suggest a higher criterion, because an old event is in doubt) and whether someone was accidentally hit while driving (which might suggest a lower criterion, because a new event is thought to be true). We thus did not have any a priori hypotheses about OCD-related changes in the criterion, but we did do an exploratory analysis testing this relationship.

The absolute values of the criterion or of signal strength cannot be identified without experimentally manipulating one or both quantities separately, since shifting them by the same amount keeps the total drift rate the same. However, we can identify bias in the criterion — changes relative to the point at which drift rate is equally likely to be towards either old or new responses — for joint settings of the signal strength and criterion. For convenience we set the criterion to 0 and signal strength to the total drift rate. For high frequency words, where there was a single *old* condition, the criterion bias was computed as the negative mean signal strength for the old and new conditions (C. N. White & Poldrack, 2014). We used a more general definition for the very low frequency condition, defining bias to be the negative of:

$$\arg\min_x |2 \cdot (1 - pnorm(x, n, s)) - (pnorm(x, o1, s) + pnorm(x, o2, s))|,$$

where $pnorm$ is the Gaussian cumulative distribution function, n, $o1$, and $o2$ are the signal strengths (drift rates) for new words, old words repeated once, and old words repeated three

times, and s is the across-trial standard deviation of drift rate. Intuitively, the quantity defined by the equation is the value of the unbiased criterion in the sense that it minimizes the difference between false positive drift rates and false negative drift rates. Because we assume the criterion is numerically 0, the bias is the negative of this amount. To obtain the relationship with Padua score, we computed the bias based on the condition drift rate effect plus the Padua drift rate effect, and subtracted from it the bias based on the condition drift rate effect alone. The resulting effects of Padua score on criterion bias are plotted in Fig. S4. There was a negative relationship between bias and Padua score for both frequency conditions in both datasets. While these relationships are hard to interpret for the reasons stated above, their consistency suggest that further study is merited.

Supplementary Discussion

Differences in boundary separation. We did not find evidence of a relationship between boundary separation in the drift-diffusion model and OCD symptoms. Given the seemingly perseverative nature of OCD, one would expect individuals with more severe symptoms to have wider decision boundaries. However, laboratory studies have revealed very mixed results on this issue. In the dot motion task, two of fives studies found wider decision boundaries (Banca et al., 2015; Erhan & Balcı, 2017), while the remaining three did not (Erhan et al., 2017; Hauser, Allen, Rees, & Dolan, 2017; Marton et al., 2019). Using the beads task, one study found that the number of draws before a decision was positively correlated with scores on the perfectionism and need for certainty subscale of the Obsessive Beliefs Questionnaire-44 (Jacoby, Abramowitz, Buck, & Fabricant, 2014), but another study found that OCD patients needed *fewer* draws (Grassi et al., 2015). In another task in which participants flipped over cards in order to decide whether the majority were one color or another, there was no difference in the number of cards flipped over by adult patients with OCD (Chamberlain et al., 2007), but pediatric patients flipped more cards over (Hauser, Moutoussis, et al., 2017). The null findings, to which our study adds, and the inconsistency of previous positive findings, underscore the fact that intuitive assumptions about how symptoms map to information processing differences are not always correct. The specific aspects of information processing which account for the more perseverative style of behavior commonly attributed to OCD have yet to be determined.

Specificity of our results. A question of obvious interest is whether our results are specific to symptoms of OCD. Previous work using a similar model-based approach found no evidence that discriminability covaried with trait anxiety (C. N. White, Ratcliff, Vasey, & McKoon, 2010; C. N. White, Ratcliff, & Vasey, 2016). However, it would be valuable to replicate both sets of findings within the same sample to measure the relative contribution of each set of symptoms. At least one study found drift rate differences during recognition related to depression (C. White, Ratcliff, Vasey, & McKoon, 2009), but it did not attempt to control for symptoms of OCD. Given the significant overlap between OCD and depression (Brown, Chorpita, Korotitsch, & Barlow, 1997; Denys, Tenney, van Megen, de Geus, & Westenberg, 2004; Pallanti, Grassi, Cantisani, Sarrecchia, & Pellegrini, 2011; Yap, Mogan, & Kyrios, 2012), establishing the specificity of these results poses significant conceptual and methodological difficulties and is not simply a matter of additional data collection. The espe-

cially high comorbidity between these two disorders is well known (for a review, see Pallanti et al., 2011). As one example, Denys et al. (2004) found that in a sample of 420 outpatients, individuals with OCD were ten times more likely to have major depressive disorder than the general population. As another example, Yap et al. (2012) found that even after explicitly dividing OCD patients into those with and without comorbid depression according to the DSM-IV, the group without comorbid depression still scored near the diagnostic cutoff for "moderate depression" on the BDI-II. And as a final example, Brown et al. (1997) found that although individuals with OCD scored lower on the depression subscale of the DASS than those with a mood disorder, they had higher scores than all other control groups (including individuals with panic disorder, social phobia, and generalized anxiety disorder) and were less than 1SD away from the mood disorder group. Not only does this pose conceptual problems for delineating the specific boundaries of each disorder, but it also poses methodological challenges for fitting statistical and cognitive models and identifying parameters as being related to one versus the other disorder. One possible path forward was illustrated by Gillan, Kosinski, Whelan, Phelps, and Daw (2016). Conducting a large factor analysis of symptoms across traditional diagnostic boundaries, the authors found separate factors which captured: 1) transdiagnostic symptoms of compulsive behavior and intrusive thought, which included symptoms of OCD, and 2) symptoms of anxiety and depression. They further showed that a specific impairment in value-based decision making was related to the first factor, but not the second. While this approach is promising, further research and other additional tools and approaches are clearly needed in order to more conclusively establish a better basis set for mapping symptoms.

Causality. It is important to note that while we framed the motivation for looking at the relationship between symptoms of OCD and recognition memory from a causal perspective, the opposite may also be true and excessive checking may result in impairments in both primary memory and metacognition (Radomsky & Alcolado, 2010). Our data are correlational and do not speak about causality specifically. However, although the ultimate causal structure remains uncertain, our results demonstrate the existence of a relationship between primary recognition memory performance and symptoms of OCD, and suggest the importance of modulating difficulty and taking a model-based approach to data analysis.

Supplementary Figures and Tables

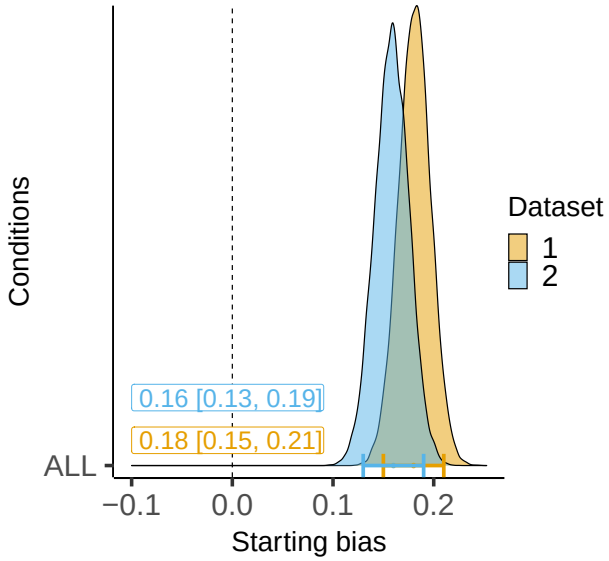

Figure S1: Starting bias (logit scale), positive numbers mean a bias towards responding "old" (the upper decision boundary).

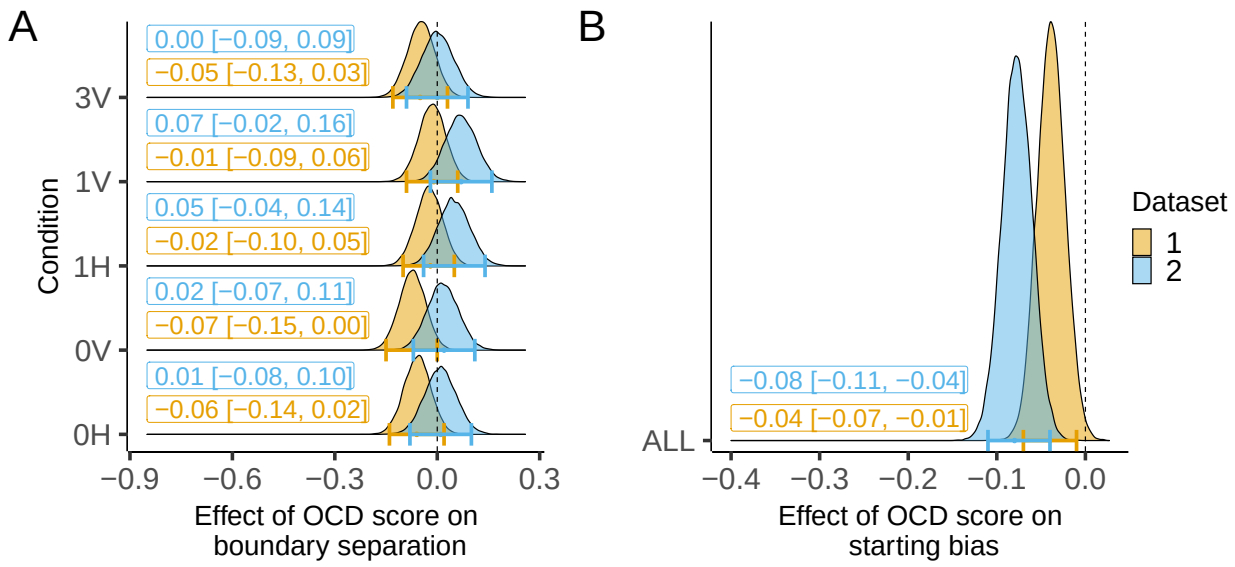

Figure S2: Effects of OCD symptoms on boundary separation and starting bias.

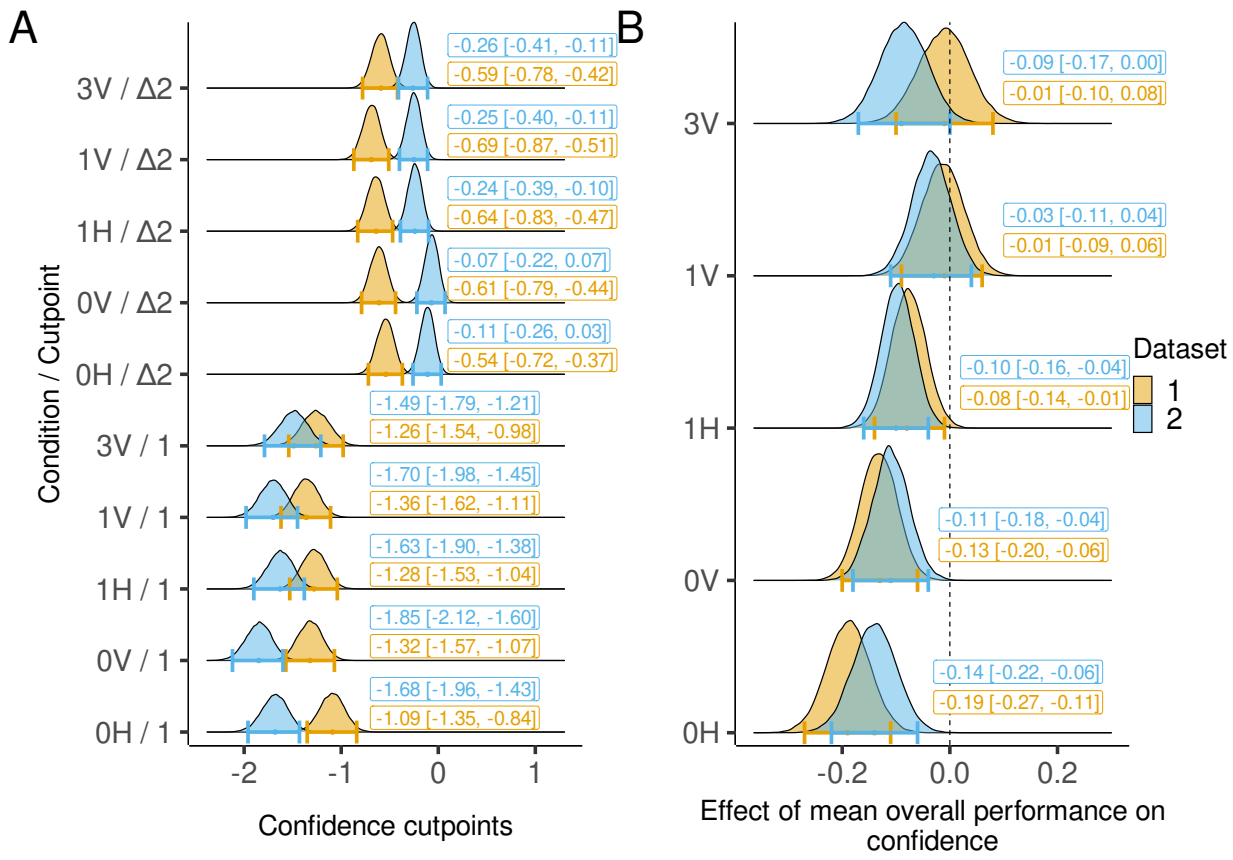

Figure S3: A. Cutpoints in the ordered probit regression model of confidence: First cutpoint and relative changes for the second cutpoint after being transformed through an exponential function (see *Methods*). B. Changes to bias in confidence due to differences in overall accuracy across trials.

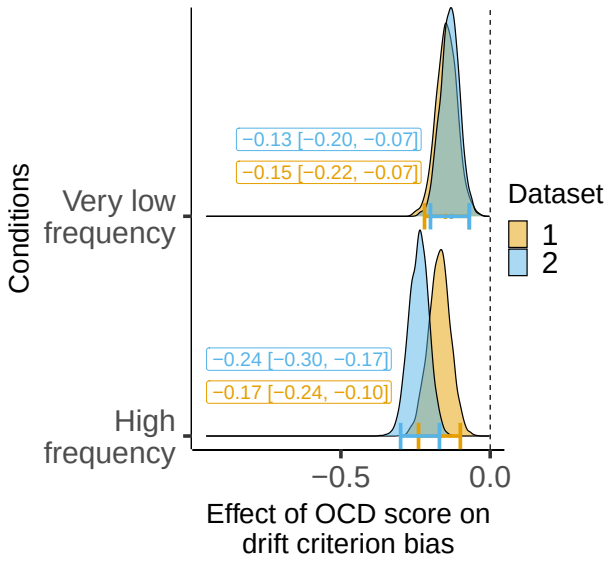

Figure S4: Effect of OCD score on bias in drift criterion: the reference relative to which drift rates are considered positive and directed toward the *old* boundary, or negative and directed toward the *new* boundary. A smaller criterion amounts to a lower memory strength requirement for judging an event to be old.

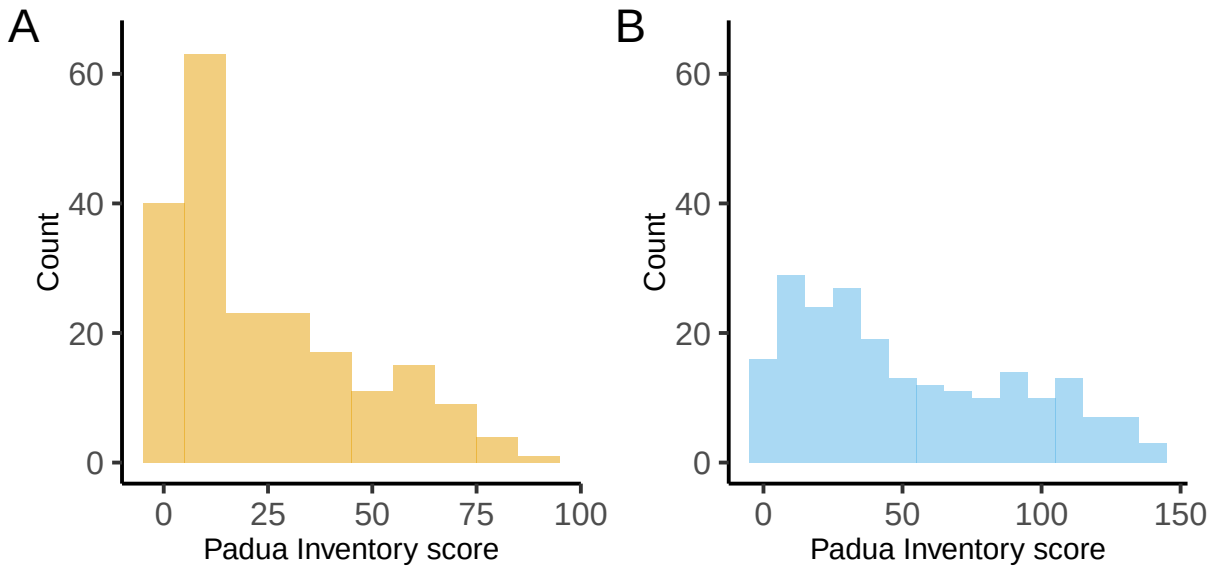

Figure S5: Padua scores in both datasets. Note that while answers are usually based on a five point Likert scale, a four point scale was used in the first experiment due to experimenter error. Part of the motivation for the second experiment was to test the replicability of effects with the scale corrected. See *Methods*. A. Dataset 1. B. Dataset 2.

Table S1: Secondary group parameters, median and 95% credible interval.

Parameter	Dataset 1	Dataset 2
Non-retrieval time, mean	0.46 [0.44, 0.48]	0.47 [0.44, 0.49]
Drift rate, sd across subjects	0.33 [0.29, 0.38]	0.40 [0.36, 0.45]
Drift rate, sd across trials	1.33 [1.28, 1.38]	1.16 [1.11, 1.21]
Boundary separation, sd	0.54 [0.48, 0.60]	0.64 [0.58, 0.72]
Starting bias, sd across subjects	0.20 [0.18, 0.23]	0.22 [0.20, 0.25]
Starting bias, sd across trials	0.41 [0.39, 0.43]	0.38 [0.36, 0.40]
Non-retrieval time, sd	0.14 [0.13, 0.16]	0.17 [0.16, 0.19]
First cutpoint, sd	1.68 [1.49, 1.91]	1.77 [1.56, 2.02]
Second cutpoint (relative to first), sd	1.12 [0.99, 1.28]	0.95 [0.84, 1.08]
Effect of being correct on confidence, sd	0.29 [0.24, 0.35]	0.32 [0.27, 0.37]

Table S2: Correlations between Padua subscales, first dataset.

	Check	Contam	Groom	Obs. Tho	Obs. Imp
Check		0.7	0.68	0.79	0.5
Contam			0.67	0.59	0.39
Groom				0.62	0.51
Obs. Tho					0.65
Obs. Imp					

Table S3: Correlations between Padua subscales, second dataset.

	Check	Contam	Groom	Obs. Tho	Obs. Imp
Check		0.78	0.86	0.85	0.71
Contam			0.76	0.7	0.56
Groom				0.77	0.69
Obs. Tho					0.87
Obs. Imp					

References

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the obsessive–compulsive inventory—revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders*, *20*(8), 1016–1035.

Banca, P., Vestergaard, M. D., Rankov, V., Baek, K., Mitchell, S., Lapa, T., ... Voon, V. (2015). Evidence accumulation in obsessive-compulsive disorder: the role of uncertainty and monetary reward on perceptual decision-making thresholds. *Neuropsychopharmacology*, *40*(5), 1192–1202.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*(1), 79–89.

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Clark, L., Robbins, T. W., & Sahakian, B. J. (2007). A neuropsychological comparison of obsessive–compulsive disorder and trichotillomania. *Neuropsychologia*, *45*(4), 654–662.

Denys, D., Tenney, N., van Megen, H. J., de Geus, F., & Westenberg, H. G. (2004). Axis I and II comorbidity in a large sample of patients with obsessive–compulsive disorder. *Journal of Affective Disorders*, *80*(2-3), 155–162.

Erhan, C., & Balcı, F. (2017). Obsessive compulsive features predict cautious decision strategies. *The Quarterly Journal of Experimental Psychology*, *70*(1), 179–190.

Erhan, C., Bulut, G. Ç., Gökçe, S., Ozbas, D., Turkakin, E., Dursun, O. B., ... Balcı, F. (2017). Disrupted latent decision processes in medication-free pediatric OCD patients. *Journal of Affective Disorders*, *207*, 32–37.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife*, *5*, e11305.

Grassi, G., Pallanti, S., Righi, L., Figee, M., Mantione, M., Denys, D., ... Stratta, P. (2015). Think twice: Impulsivity and decision making in obsessive–compulsive disorder. *Journal of Behavioral Addictions*, *4*(4), 263–272.

Hauser, T. U., Allen, M., Rees, G., & Dolan, R. J. (2017). Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Scientific Reports*, *7*, 6614.

Hauser, T. U., Moutoussis, M., Iannaccone, R., Brem, S., Walitza, S., Drechsler, R., ... Dolan, R. J. (2017). Increased decision thresholds enhance information gathering performance in juvenile obsessive-compulsive disorder (OCD). *PLOS Computational Biology*, *13*(4), e1005440.

Jacoby, R. J., Abramowitz, J. S., Buck, B. E., & Fabricant, L. E. (2014). How is the Beads Task related to intolerance of uncertainty in anxiety disorders? *Journal of Anxiety Disorders*, *28*(6), 495–503.

Marton, T., Samuels, J., Nestadt, P., Krasnow, J., Wang, Y., Shuler, M., ... Nestadt, G. (2019). Validating a dimension of doubt in decision-making: A proposed endophenotype for obsessive-compulsive disorder. *PLOS ONE*, *14*(6).

Pallanti, S., Grassi, G., Cantisani, A., Sarrecchia, E., & Pellegrini, M. (2011). Obsessive–compulsive disorder comorbidity: clinical assessment and therapeutic implications. *Frontiers in Psychiatry*, *2*(70).

Radomsky, A. S., & Alcolado, G. M. (2010). Don't even think about checking: Mental checking causes memory distrust. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(4), 345–351.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Wells, A. (2002). *Emotional disorders and metacognition: innovative cognitive therapy*. John Wiley & Sons.

White, C., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion*, *23*(1), 181–205.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 385–398.

White, C. N., Ratcliff, R., & Vasey, M. W. (2016). Anxiety-related threat bias in recognition memory: the moderating effect of list composition and semantic-similarity effects. *Cognition and Emotion*, *30*(8), 1446–1460.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, *54*(1), 39–52.

Yap, K., Mogan, C., & Kyrios, M. (2012). Obsessive-compulsive disorder and comorbid depression: the role of OCD-related and non-specific factors. *Journal of Anxiety Disorders*, *26*(5), 565–573.