Metacognition and the effect of incentive motivation in two compulsive disorders: gambling disorder and obsessive-compulsive disorder

Authors:

Monja Hoven[1], Nina S. de Boer[2] , Anna E. Goudriaan[1,3], Damiaan Denys[1], Mael Lebreton[4,5,6], Ruth J. van Holst[1*], Judy Luigjes[1*]

*: Shared last authors

Affiliations:

[1]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
[2]Department of Philosophy, Radboud University, Nijmegen, The Netherlands
[3]Arkin and Jellinek, Mental Health Care, and Amsterdam Institute for Addiction Research, Amsterdam, The Netherlands
[4]Paris School of Economics, Paris, France
[5]Swiss Center for Affective Science, University of Geneva, Switzerland
[6]Laboratory for Behavioral Neurology and Imaging of Cognition, Department of Fundamental Neurosciences, University of Geneva, Switzerland

Funding: Data collection for this work was funded by two independent personal Amsterdam Brain and Cognition (ABC) Talent grants to JL and RvH, and a NWO Veni Fellowship (grant 451-15-015) granted to ML. ML is supported by a Swiss National Fund Ambizione Grant (PZ00P3_174127) and an ERC Starting Grant (ERC-StG-948671), JL is supported by a NWO VENI Fellowship grant (916-18-119).

Nr. of words: 7183

Keywords: confidence, metacognition, vmPFC, obsessive-compulsive disorder, gambling disorder, compulsivity

Abstract:

Compulsivity is a common phenotype amongst various psychiatric disorders, such as obsessive-compulsive disorder (OCD) and gambling disorder (GD). Deficiencies in metacognition, such as the inability to properly estimates ones' own performance via well-calibrated confidence judgments could contribute to pathological decision-making in these psychiatric disorders. Earlier research has indeed suggested that OCD and GD patients reside at opposite ends of the confidence spectrum, with OCD patients exhibiting underconfidence, and GD patients exhibiting overconfidence. Recently, several studies established that motivational states (e.g. monetary incentives) influence metacognition, with gain (respectively loss) prospects increasing (respectively decreasing) confidence judgments. Here, we reasoned that the OCD and GD symptomatology might correspond to an exacerbation of this interaction between metacognition and motivational states. We hypothesized GD's overconfidence to be exaggerated during gain prospects, while OCD's underconfidence to be worsened in loss context, which we expected to see represented in ventromedial prefrontal cortex (VMPFC) blood-oxygen-level-dependent (BOLD) activity. We tested those hypotheses in a task-based functional magnetic resonance imaging (fMRI) design. Our initial analyses showed increased confidence levels for GD versus OCD patients, that could partly be explained by sex and IQ. Although our primary analyses did not support the hypothesized interaction between incentive motivation and groups, exploratory analyses did show increased confidence in GD patients specifically in gain context. fMRI analyses confirmed a central role for VMPFC in the processing of confidence and incentive motivation, but with no differences between the clinical samples.

Introduction

Compulsive behaviors are defined as “repetitive acts that are characterized by the feeling that one ‘has to’ perform them while one is aware that these acts are not in line with one’s overall goal” (Luigjes et al., 2019). Various psychiatric disorders are associated with compulsivity, of which obsessive-compulsive disorder (OCD) is likely the most typical (Stein, 2002), but is also seen in addictive disorders such as gambling disorder (GD) (van Timmeren et al., 2018). Both these disorders are characterized by performing compulsive behaviors and loss of control over those behaviors, albeit originating from distinct motivations, serving different purposes and relating to distinct symptomatology (Chamberlain et al., 2005; Figee et al., 2016). Where OCD patients mostly perform compulsions (i.e. washing, checking, counting) to decrease their feelings of anxiety evoked by obsessions and are regarded to be loss avoidant (Kaufmann et al., 2013; Stein, 2002), GD patients typically compulsively gamble following their sensitivity for rewards (Grant et al., 2016; Sescousse et al., 2013; Van Holst et al., 2012) and due to craving (Van Holst et al., 2012). Hence, compulsivity seems to be a common phenotype in otherwise symptomatically different disorders.

Dysfunctions in metacognition could explain both overlap and distinct features of compulsive behaviors. Metacognition is the ability to monitor, reflect upon and think about our own behavior (Fleming et al., 2012). While metacognition is a broad umbrella term under which many different cognitive computations gather, some specific computations can be formally defined. One of those is the judgment of confidence, which is defined as the subjective estimate of the probability of being correct about a choice (Pouget et al., 2016). Confidence plays a key role in decision-making and learning (Fleming et al., 2012; Meyniel et al., 2015; Pouget et al., 2016), and therefore in steering our future behavior (Folke et al., 2017; Samaha et al., 2019). It is crucial for behavioral control that one’s confidence is in line with reality. Nonetheless, persistent discrepancies between actual behavior (e.g. choice accuracy) and confidence in that behavior (subjective estimate of accuracy) have been consistently described, which could contribute to pathological (compulsive) decision-making

as seen in various psychiatric disorders (Hoven et al., 2019). Clinical presentations of OCD and GD indeed suggest confidence abnormalities in opposite direction, under- and overconfidence, respectively, which could both promote detrimental decision-making, such as checking behavior and compulsive gambling (Fortune & Goodie, 2012; Goodie & Fortune, 2013; Nestadt et al., 2016; Samuels et al., 2017). We recently reviewed the literature and found that both people with subclinical and clinical OCD consistently showed a decrease in confidence level, which was especially found in OCD-symptom contexts (Hoven et al., 2019). Also, confidence abnormalities and pathological compulsivity were found to be closely coupled in OCD: engaging in compulsive behaviors increasingly lowered subjects' confidence levels and, simultaneously, experimentally decreasing confidence reinforced subjects' compulsive behaviors. Oppositely, in pathological gamblers, there was evidence for overconfidence in rewarding gambling contexts (Hoven et al., 2019), which was also related to symptom severity (Goodie, 2005; Lakey et al., 2007). In sum, GD and OCD patients seem to function at opposite sides of the confidence continuum, respectively over- and under-estimating their performance, which could explain how opposite traits may underlie similar pathological behavior (i.e. compulsive behavior).

Theoretical models of confidence formation suggest that confidence builds – at least partly - on the integration of noisy evidence used for decision-making (Fleming & Daw, 2017; Sanders et al., 2016). A resulting signature of confidence is its statistical dependence on an interaction of accuracy and perceptual evidence, which is typically illustrated as an 'X-pattern' where confidence increases/decreases with increasing evidence for correct/incorrect decisions, respectively (Fleming & Daw, 2017; Sanders et al., 2016). A recent study showed that trait compulsivity is related to deficits in using perceptual evidence to inform confidence in a healthy population (Seow & Gillan, 2020). It remains unknown, however, if GD and OCD patients likewise show aberrant integration of evidence in confidence signals.

What is known about both OCD and GD, however, is that patients with those pathologies show deficits in reward and motivation processes and accompanying dysregulated neural

reward circuitries (Admon et al., 2012; Choi et al., 2012; Clark et al., 2019; Figee et al., 2011; Jung et al., 2011; Meng et al., 2014; Romanczuk-Seiferth et al., 2015). This is particularly relevant to the question of how confidence might contribute to those pathologies' symptoms, as an increasing number of studies show that affective and motivational states, such as mood and arousal, can influence confidence (Allen et al., 2016; Koellinger & Treffers, 2015; Massoni, 2014). Recently, we demonstrated that monetary incentives bias confidence judgments in healthy individuals, where prospects of gain (respectively loss) increase (respectively decrease) confidence, whilst performance levels remained unaffected in both perceptual and reinforcement-learning contexts (Hoven et al., 2020; Lebreton et al., 2018; Lebreton, Bacily, et al., 2019; Ting et al., 2020).

We therefore reasoned that, since motivational states (i.e. incentive contexts) influence confidence judgments, deficits in reward and motivation processes could percolate to confidence judgments. In other words, an interaction between motivation and confidence could cause or fuel the compulsive behaviors observed in GD and OCD. On the one hand, prospects of high monetary incentives could exaggerate general overconfidence in GD patients. On the other hand, as harm avoidance is considered one of the core motivations of compulsive behavior in OCD (Bey et al., 2017, 2020; Summerfeldt et al., 2014), this could drive an overall decreased confidence, which is specifically exaggerated in negative value contexts in OCD patients.

On the neurobiological side, a growing number of functional magnetic resonance imaging (fMRI) studies have associated metacognitive processes with activity in the frontal-parietal network (Allen et al., 2017; Baird et al., 2013; Fleming et al., 2010; Hilgenstock et al., 2014; Vaccaro & Fleming, 2018) , and blood-oxygen-level-dependent (BOLD) activity in the dorsomedial prefrontal cortex (dmPFC), insula and dorsal anterior cingulate cortex (dACC) has been negatively associated to confidence, suggesting a role for these areas in representing uncertainty-related variables (Fleming et al., 2018; Molenberghs et al., 2016; Morales et al., 2018; Rouault & Fleming, 2020; Shenhav et al., 2016). Interestingly, recent

studies have also found activity in the ventral striatum and the median wall of the prefrontal cortex – i.e. the ventromedial PFC (vmPFC) and perigenual anterior cingulate cortex (pgACC) - to be positively associated with confidence (Bang & Fleming, 2018; De Martino et al., 2013; Gherman & Philiastides, 2018; Hebart et al., 2016; Lebreton et al., 2015; Rouault, Mcwilliams, et al., 2018; Rouault & Fleming, 2020). Importantly, this latter network has been previously associated with motivation and value-based processes (Bartra et al., 2013; Haber & Behrens, 2014; Haber & Knutson, 2009; Lopez-Persem et al., 2020). Actually, both confidence judgments and value information seem to be automatically integrated into the activity in the vmPFC during decision making (Gherman & Philiastides, 2015; Lebreton et al., 2009, 2015; Lopez-Persem et al., 2020; Shapiro & Grafton, 2020). Yet, little is known about if and how the behavioral interaction observed between motivation and confidence can be explained by their shared association with the vmPFC. In an attempt to answer this question, we recently reported an important interaction between motivational and metacognitive signals in the vmPFC: VMPFC confidence signals were observed in trials with gain prospects, but disrupted in trials with no – or negative (loss) monetary prospects (Hoven et al., 2020). This suggest that VMPFC has a key role in mediating the relation between motivation and metacognition.

In the present study we investigate metacognitive ability and its interaction with incentive motivation in two disorders with phenomenological different types of compulsive behaviors central to their pathology: OCD and GD. First, we tested how OCD patients, GD patients and healthy controls (HCs) differed at the behavioral level. Then, using fMRI, we investigated the neurobiological foundations of incentive motivation and confidence encoding in GD, OCD and HCs, in an attempt to shed light on the brain regions involved in dysfunctional confidence computation.

Methods

Ethics

All experimental procedures were approved by the Medical Ethics Committee of the Academic Medical Center, University of Amsterdam. All subjects provided written informed consent.

Participants

We recruited a total of 31 GD patients, 29 OCD patients and 55 HCs for this study. Of our HC sample of 55 subjects, 25 subjects were included in our earlier work (Hoven et al., 2020). All subjects were aged between 18 and 65 years old. HCs were recruited through online advertisements and from our participant database. GD patients were recruited from a local addiction center (Jellinek Addiction Treatment Center Amsterdam) and were recently diagnosed with GD. OCD patients were recruited through the department of psychiatry at the Academic Medical Center in Amsterdam and were diagnosed with OCD. All subjects underwent screening with the MINI structured psychiatric interview to confirm the absence of any other psychiatric disorder (Sheehan et al., 1998). OCD symptom severity was measured using the Yale-Brown Obsessive Compulsive Scale (YBOCS) (Goodman et al., 1989), and GD symptom severity was measured using the Problem Gambling Severity Index (PGSI) (Ferris & Wynne, 2001). Anxiety symptoms were assessed using the Hamilton Anxiety Rating Scale (HAMA) (Hamilton, 1959) and depression symptoms using the Hamilton Rating Scale for Depression (HDRS) (Hamilton, 1960).

Exclusion criteria

The exclusion criteria included having a diagnosis of major depressive disorder, bipolar disorder, psychotic disorders, substance-use disorders, using tricyclic antidepressants or

antipsychotics, having any contraindications for MRI, and having a history of or current treatment for neurological disorders, major physical disorders or brain trauma.

Moreover, session-level behavioral and fMRI data were excluded when task accuracy was below 50% or when subjects did not show sufficient variation in their confidence reports (standard deviation of <5 confidence points), and session-level fMRI data was additionally excluded when participants displayed more than 3.5 mm head movement in any direction. Overall, for the behavioral analyses, this led to the full exclusion of four GD patients and one OCD patient, as well as one out of two session exclusions for four GD patients, two OCD patients and two HCs. For the fMRI analyses, three additional GD patients, one OCD patient and two HCs were fully excluded, as well as one out of two session exclusions for three additional GD patients and one OCD patient. Thus, we finally included 27 GD patients, 28 OCD patients and 55 HCs for the behavioral analyses, of which four, two and two participants contributed only one of two task sessions, respectively. For the fMRI analyses we included 24 GD patients, 27 OCD patients and 53 HCs, of which seven, three and two participants contributed only one of two task sessions, respectively.

Demographics

We analyzed whether age, sex, IQ, Y-BOCS, PGSI, HAMA and HDRS score differed between the three groups using ANOVAs for all variables but sex, which was assessed using a Chi-square test. When appropriate, two-sample t-tests were executed post-hoc.

Experimental Design and Study Procedure

We used a similar experimental design and study procedure as described in (Hoven et al., 2020). For details on the experimental design and study procedure, see Hoven et al. (2020) and **Figure 1**. In sum, subjects performed a simple perceptual decision-making task, with a

2-alternative forced choice of contrast discrimination followed by a confidence judgment. In each trial, participants could either win (gain context) or lose (loss context) points, or not (neutral context), conditional on the accuracy of the choice in that trial. Importantly, this incentivization was administered after the choice moment, but before the confidence rating. The task was implemented using MATLAB® (MathWorks Inc., Sherborn, MA, USA) and the COGENT toolbox.

After demographic and clinical interviews, all participants performed an initial calibration session (consisting of 144 trials) to tailor the difficulty level of the task to each individual. This was done to keep average performance similar across individuals. Following, all subjects performed two fMRI sessions, each consisting of 72 trials (24 per incentive condition), presented in a random order.

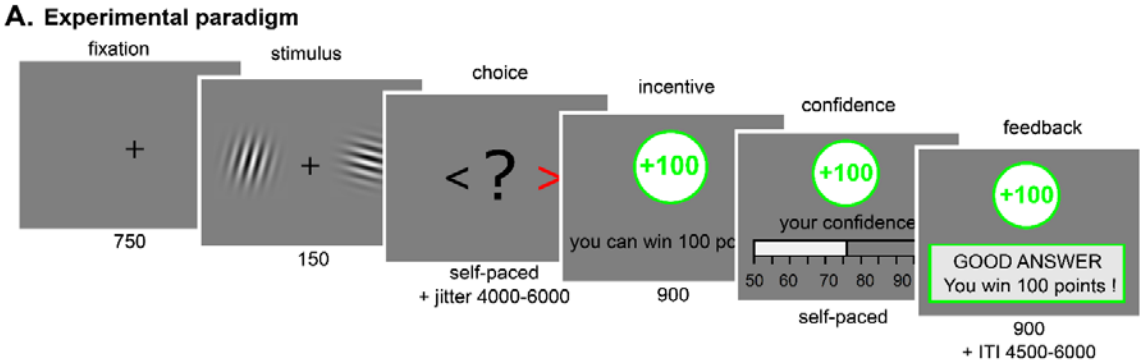

Figure 1 | Experimental paradigm. Participants viewed two Gabor patches on both sides of the screen (150 ms) and then chose which had the highest contrast (left/right, self-paced) (for more information, see Hoven et al., 2020). After a jitter of a random interval between 4500 to 6000 ms, the incentive was shown (900 ms; green frame for win trials, grey frame for neutral trials, red frame for loss trials). Afterwards, participants were asked to report their confidence in their choice on a rating scale ranging from 50% to 100% with steps op 5%. The initial position of the cursor was randomized between 65% and 85%. Finally, subjects received feedback. The inter trial interval (ITI) had a random duration between 4500 and 6000 ms. The calibration session only consisted of Gabor discrimination, without confidence rating, incentives or feedback and was used to adjust difficulty so that every individual reached a performance of 70%.

Behavioral Measures

We extracted trial-by-trial experimental factors: incentive condition and evidence and behavioral measures: accuracy, confidence ratings, reaction times. In addition, we computed an extra *latent* variable: early certainty.

The early certainty variable was computed in order to analyze BOLD activity at choice moment, when the brain encodes a confidence signal that is not yet biased by incentives. This was done by making a trial-by-trial prediction of early certainty based on stimulus features (reaction times, evidence and accuracy) at choice moment. This resulted in an early certainty signal that was highly correlated with confidence, but showed no statistical relationship with incentives (see **Supplementary Materials**). For more details, see Hoven et al. (2020).

Behavioral Analyses

All analyses were performed in the R environment (RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA). We used linear mixed effects models (LMEMs) as implemented in the lmer function from the lme4 and afex packages (Bates et al., 2015; Singmann et al., 2015). To determine p-values for the fixed effects, we performed Type 3 F tests with Satterthwaite approximation for degrees of freedom as implemented in the afex package. When relevant, we used the ‘emmeans’ package to perform post-hoc tests that were corrected for multiple comparisons using Tukey’s method (Lenth et al., 2018).

To answer our main research questions, we built several LMEMs and performed a model selection procedure. We iteratively built LMEMs and compared those by assessing model fit by using Chi-square tests on the log-likelihood values, and by comparing of the AIC and BIC model values. We started with a basic model with fixed effects of incentive, group and their interaction on confidence, together with a random subject intercept and slope of incentive per

subject. Model predictors of accuracy and evidence, together with their interaction and the interaction with group were added whenever it significantly improved model fit. See **Table 1** for the model comparison results. The final model (henceforth termed Model 1) consisted of fixed effects of incentive, group, accuracy and evidence (z-scored) and interactions between incentive and group, as well as two-way and three-way interactions between evidence, accuracy and group. All models included trial-by-trial data, and a random subject intercept as well as a random slope of incentive per subject.

In order to confirm that the incentive condition or group did not influence accuracy or reaction time, we modelled additional LMEMs with performance and reaction time as dependent variables (Model 2, Model 3).

Lastly, in order to control for the differences in the distribution of the demographic variables IQ and sex between groups, we added IQ (z-scored) and sex as fixed effects to our original Model 1 (Model 4). Model fit was assessed and compared using Chi-square tests on log-likelihood values. Additional control analyses on the properties of confidence and early certainty are reported in the **Supplementary Materials**.

Due to a technical bug, our design was not fully balanced as the level of perceptual evidence was not equal across the incentive conditions. ANOVA and post-hoc testing indeed showed that evidence was highest in neutral condition, followed by gain, and loss. There were no group differences, nor an interaction between group and incentive. These effects cannot account for any group differences we find in our data, since evidence did not differ between groups. Importantly, the evidence differences did not affect performance, since performance is equal across conditions. See **Supplementary Results** for more details.

fMRI acquisition and preprocessing

For details on fMRI acquisition see Hoven et al (2020). All our analyses were performed using MATLAB with SPM12 software (Wellcome Department of Cognitive Neurology,

London, UK). Raw multi-echo functional scans were weighed and combined into 570 single volumes per scan session (Poser et al., 2006), using the first 30 dummy scans to calculate the optimal weighting of echo times for each voxel by applying a PAID-weight algorithm. During the combining process, realignment was performed on the functional data by using linear interpolation to the first volume. Subsequently, the functional images were co-registered, segmented for normalization to MNI space and smoothed. To reduce motion-related artifacts, the Art-Repair toolbox (Mazaika et al., 2009) was used to detect large volume-to-volume movement and repair outlier volumes.

fMRI analyses

All fMRI analyses were conducted using SPM12. Critically, our design allowed us to distinguish between our two timepoints of interest: 1) the moment of stimulus presentation and choice in which implicit (un)certainty about the choice is formed, and 2) the moment of incentive presentation and confidence rating, in which the value of incentives and the confidence rating are encoded. We built a general linear model (GLM 1) that was estimated on subject-level with these two moments of interest: the moment of choice (i.e. presentation of the Gabor patches) and the moment of incentive presentation/confidence rating. We chose to analyze the incentive presentation and confidence rating as a single timepoint since the rating moment followed the presentation of the incentive after 900 ms. We also included a regressor for the moment of feedback to explain variance in neural responses related to feedback on accuracy and value that was not related to the decision-making process, but this regressor was not of interest for the current analyses. All whole-brain activation maps were thresholded using family-wise error correction (FWE) at cluster level ($P_{FWE_clu} < 0.05$), with a voxel cluster-defining threshold of $p<.001$ uncorrected.

GLM 1 consisted of three regressors for each timepoint: ‘choice’, ‘incentive/rating’ and ‘feedback’, to which parametric modulators (pmods) were added. All regressors were

specified as stick functions time-locked to the onset of the respective events. The choice regressor was modulated by two pmods: early certainty (z-scored on subject level) and button press (left or right) to control for motor-related activation. The incentive/rating regressor was modulated by two pmods: incentive value ([-1,0,1]) and confidence rating (z-scored on subject level). The feedback regressor was additionally modulated by a pmod representing choice accuracy. Our contrasts of interest included: (1) choice moment modulated by early certainty, (2) incentive/rating moment modulated by incentive value and (3) incentive/rating moment modulated by confidence rating.

In order to study the interaction between incentive motivation and metacognitive ability on the neurobiological level we leveraged the factorial design of our task to build GLM 2. We used this GLM 2 to explicate the effect of incentive motivation on both the integration of evidence at choice moment, as well as on confidence formation, and compare those between groups. For this model we built a regressor for each of two time points (choice moment and incentive/rating moment) and three incentive conditions, as well as a single regressor at feedback moment, resulting in seven regressors. All regressors at choice moment were modulated by a pmod of button press (left/right) and *signed evidence:* a variable that signifies the interaction between evidence and accuracy. Signed evidence was calculated as the absolute value of evidence in case of correct answers and the negative absolute evidence (i.e. –abs(evidence)) in case of incorrect answers. All regressors at rating moment were modulated by a pmod of confidence, and the feedback regressor was modulated by a pmod of accuracy. Thus, for all these events we could examine both baseline activity and regression slopes relating to their respective pmod. Since the results by Hoven et al., 2020 suggested that the VMPFC plays an important role in the interaction between motivation and metacognition, we created a functional region of interest (ROI) that represented the confidence-related activity in the VMPFC cluster from our GLM 1 across groups results (see **Figure 4C**, **Table 5**). We then extracted individual *t*-statistics within this ROI (i.e. normalized beta estimates, see Lebreton et al. (2019)) from our contrasts of interest and performed one-

sample t-tests against 0 to check for positive or negative activation patterns. Then, we compared them between incentive conditions, groups, and studied their interactions using mixed ANOVAs implemented in the afex package. When appropriate, we performed post-hoc testing using the emmeans package, correcting for multiple comparisons using Tukey's method.

For both GLMs pmods were not orthogonalized and thus competed to explain variance. We included six motion parameters as nuisance regressors. Regressors were modeled separately for each scanning session and constants were included to account for between-session differences in mean activation. All events were modeled by convolving a series of delta functions with the canonical hemodynamic response function (HRF) at the onset of each event and were linearly regressed onto the functional BOLD-response signal. Low frequency noise was filtered with a high pass filter with a cut off of 128 seconds. We controlled for the number of sessions while making the first-level contrasts. All contrasts were computed at subject level and then taken to group level analyses. For GLM 1 we assessed group differences by performing a one-way ANOVA to our contrasts of interest, using an F-contrast test to test for any group differences (i.e. [1 -1 0; 0 1 -1]). In addition, to gain a complete picture of areas involved in our contrasts of interest, we grouped all subjects together and performed one-sample t-tests against 0.

Model	Model notation	AIC	BIC	Comparison	X2	P-value	Winning model
A	Confidence ~ Incentive * Group + (Incentive|Subject)	122919	123041				
B	Confidence ~ Incentive * Group + Accuracy + (Incentive|Subject)	122273	122402	A vs. B	648.59	< 2.2e-16	**B**
C	Confidence ~ Incentive * Group + Accuracy + Evidence + (Incentive|Subject)	122004	122141	B vs. C	271.00	< 2.2e-16	**C**
D	Confidence ~ Incentive * Group + Accuracy*Evidence + (Incentive|Subject)	121791	121936	C vs. D	214.53	< 2.2e-16	**D**
E	Confidence ~ Incentive * Group + Accuracy*Evidence*Group + (Incentive|Subject)	121751	121942	D vs. E	52.141	1.747e-09	**E**
F	Confidence ~ Incentive * Group + Sex + IQ (Incentive|Subject)	121752	121958	E vs. F	2.7018	0.259	**E**

Table 1 | Model descriptions and comparison. Shown here are the model notations of all models with their respective Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, as well as model comparison outcomes with corresponding $\chi 2$ and P-values, resulting in the winning model 'E', which is referred to as Model 1 in the manuscript.

Results

Demographics

Age was not significantly different between the three groups ($F_{2,107} = 0.253$, $p > 0.75$), but IQ was, ($F_{2,107} = 3.222$, $p = 0.0438$). Post-hoc t-tests showed that HC subjects had a significantly higher IQ score than GD patients ($t = 2.53$, $p=0.014$). As expected, Y-BOCS scores and PGSI scores differed significantly between groups ($F_{2,107} = 322.2$, $p<.001$;$F_{2,107} = 380.5$, $p<.001$, respectively), with OCD patients having higher Y-BOCS scores than HCs ($t = -16.97$, $p<.001$) and GD patients ($t = -36.67$, $p<.001$), and GD patients having higher PGSI scores than HCs ($t = -15.99$, $p<.001$) and OCD patients ($t = -14.32$, $p<.001$). HAMA scores were significantly different between groups ($F_{2,107} = 48.02$, $p<.001$), post-hoc tests revealed higher

HAMA scores for OCD patients than HCs (t = -8.50, p<.001) and GD patients (t = 4.58, p<.001), and higher HAMA scores for GD patients compared to HCs (t = -2.44, p=.002). HDRS scores were significantly different between groups ($F_{2,107}$ = 24.97, p<.001), with higher scores for OCD versus HC (t = -7.76, p<.001), and higher scores for GD versus HC (t=-3.03, p=.005). Lastly, using a Chi-square test we found a significant difference in sex distribution between the groups (X = 14.483, df = 2, p<.001), with mostly men in the GD group, and relatively more women in the OCD group (Table 2). This corresponds to the natural distribution observed in epidemiological studies for OCD and GD, showing higher prevalence of GD amongst men, and a slightly higher prevalence of OCD in women (Black & Shaw, 2019; Calado & Griffiths, 2016; Mathes et al., 2019; Ruscio et al., 2008).

	HC	**GD**	**OCD**	**Statistics**
Age	33.51 +- 12.32	33.22 +- 10.40	31.93 +- 8.21	$F_{2,107}$ = 0.25, p = 0.777
IQ*	91.18 +- 10.96	85.22 +- 9.53	89.54 +- 8.32	$F_{2,107}$ = 3.22, p = 0.0438
Y-BOCS***	0.25 +- 1.76	1.19 +- 2.60	20.36 +- 6.15	$F_{2,107}$ = 322.2, p<.001
PGSI***	0.05 +- 0.40	14.85 +- 4.80	0.64 +- 1.91	$F_{2,107}$ = 380.5, p<.001
HAMA***	1.09 +- 1.97	3.93 +- 5.88	11.43 +- 6.28	$F_{2,107}$ = 48.02, p<.001
HDRS***	1.31 +- 2.31	5.07 +- 6.24	7.71 +- 4.04	$F_{2,107}$ = 24.97, p<.001
Sex (m/f)***	33 / 22	24 / 3	11 / 17	$X^2(2)$ = 14.483, p<.001

Table 2 | Demographics: Means +- standard deviations of various demographic variables are shown per group, for sex counts are displayed. Statistics for group comparisons are shown, including F and X^2 statistics, degrees of freedom and p-values. IQ= estimated Intelligence Quotient, GD = gambling disorder, HAMA = Hamilton Anxiety Rating Scale, HC = healthy control, HDRS = Hamilton Depression Rating Scale, OCD = obsessive-compulsive disorder PGSI = Problem Gamblers Severity Index, Y-BOCS = Yale-Brown Obsessive Compulsive Scale. *p<.05, ***p<.001

Behavioral Results

To start, we answered our first three questions: (1)are there group differences in confidence, (2) what is the influence of incentive motivation on confidence, and (3) how is evidence

integrated into confidence in the groups? Model 1 showed a main effects of group ($F_{2,112}$ = 4.7910, p=.01) and incentive ($F_{2,112}$ = 20.9371, p<.001) on confidence (**Figure 2, Supplementary Table 3**). We also found a main effect of accuracy ($F_{1,15107}$ = 608.8906, p<0.001), with subjects showing higher confidence for correct answers. Moreover, there was a significant two-way interaction of group and evidence ($F_{2,15099}$ = 3.5094, p = 0.02994). As expected, we also found a significant interaction between accuracy and evidence, replicating the 'X-pattern' signature of evidence integration where confidence increases with increasing evidence when correct, and vice versa ($F_{1,15097}$=185.3245, p<0.001) (Sanders et al., 2016). Interestingly, the evidence integration effect differed per group, as signaled by a significant three-way interaction between accuracy, evidence and group ($F_{2,15094}$ = 3.0533, p=0.04723) (**Figure 3, Supplementary Table 3**). Lastly, the interaction between incentive and group revealed a trend towards an effect ($F_{4,112}$ = 2.2821, p=0.06487).

Post-hoc tests indicated a significantly higher confidence in GD patients versus OCD patients (GD-OCD = 6.38 +- 2.12, Z-ratio = 3.014, p = 0.0073), and a trend towards higher confidence in GD compared to HC subjects (GD-HC = 4.30 +- 1.84, Z-ratio = 2.333, p = 0.0513), whereas OCD patients did not differ from HC subjects. Moreover, we replicated the parametric effect of incentive value on confidence (loss-neutral =-1.80 +- 0.429, Z-ratio = -4.192, p<0.001; loss-gain =-3.14 +- 0.486, Z-ratio = -6.460, p<0.001; neutral-gain = -1.34 +- 0.363, Z-ratio = -3.683, p<0.001). With regards to the three-way interaction, we compared the groups on the slopes of evidence integration in confidence separately for correct and incorrect trials using the emtrends() function, and found that the slope for evidence integration into confidence was less steep for correct answers in GD patients compared to both HCs (GD - HC = -1.712 +- 0.283, Z-ratio = -6.057, p<0.001) and OCD patients (GD - OCD = -2.110 +- 0.357, Z-ratio = -5.912, p<0.001). This indicates that GD patients' confidence ratings were less influenced by the perceptual evidence when they made a correct choice. No differences between OCD patients and HC were found regarding evidence integration effects. Exploratory post-hoc analyses on the group*incentive

interaction effect showed that, especially in context of possible gains, GD patients were more confident than OCD patients (GD - OCD = 8.12 +- 2.24, Z-ratio = 3.621, $p < 0.001$) and HC subjects (GD - HC = 5.83 +- 1.95, Z-ratio = 2.989, p=0.0079), whereas no differences were observed between HC and OCD patients in any incentive condition (**Table 3**).

We also estimated Model 2 and 3, with accuracy and reaction time as dependent variables (**Table 4**). These models revealed no effect of group, incentive or an interaction effect on accuracy or reaction time, as expected from our design (where incentives are revealed after choices). Importantly, these control analyses confirm that accuracy and response times cannot confound any effect of incentives that we found on confidence.

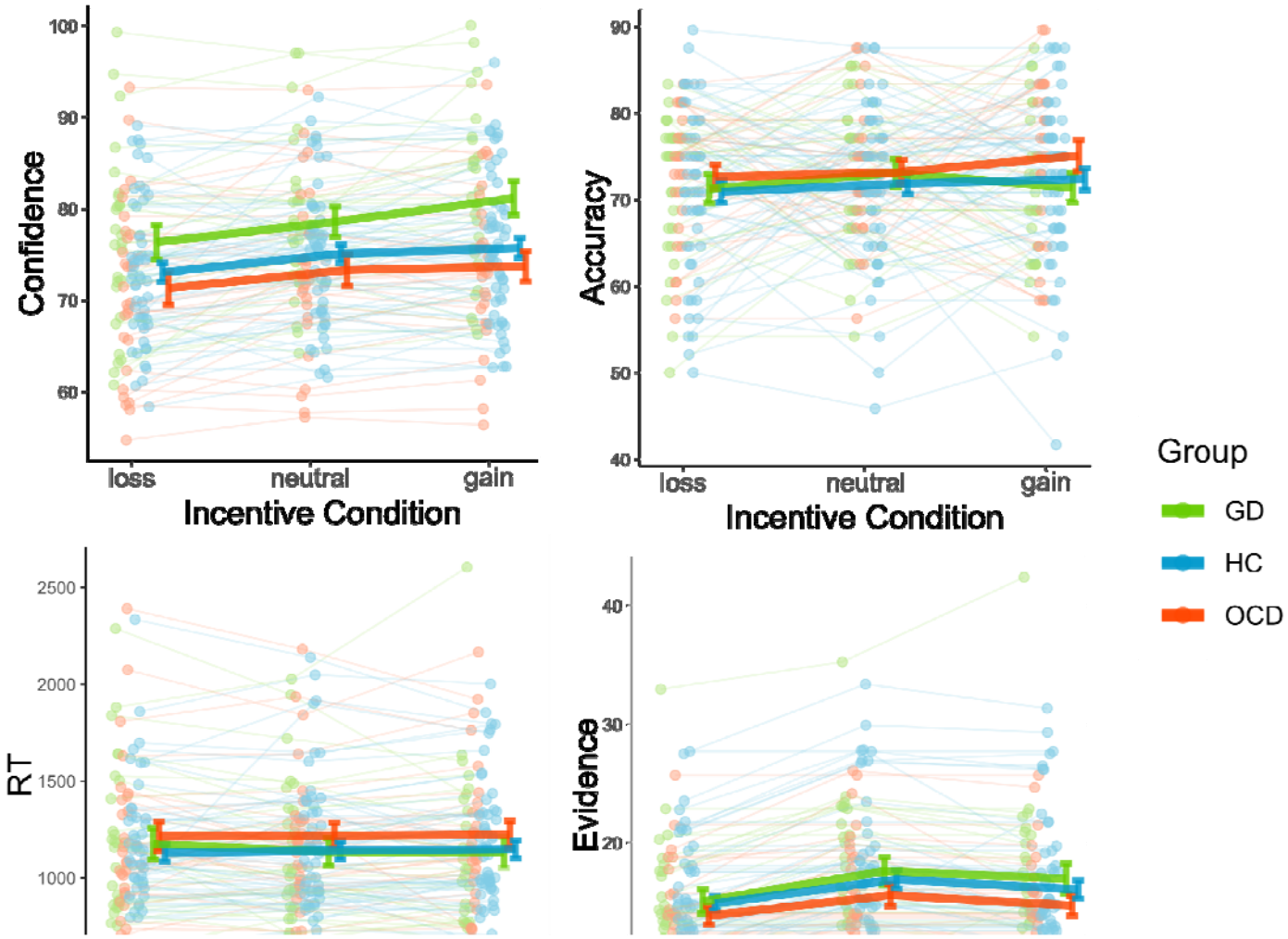

Figure 2 | Behavioral results. Individual-averaged confidence, accuracy, reaction times and evidence as a function of incentive condition (loss, neutral and gain) per group. Green dots and lines represent gambling disorder patients, blue dots and lines represent healthy controls and red dots and lines represent obsessive-compulsive disorder patients. Dots represent individuals, and lines highlight within subject variation across conditions. Error bars represent sample mean ± SEM per group. GD = gambling disorder, HC = healthy control, OCD = obsessive-compulsive disorder

Model 1	Confidence
Incentive	F(2.00, 112.34) = 20.94, p < .001
Group	F(2.00, 112.51) = 4.79, p = .010
Accuracy	F(1.00, 15107.05) = 608.89, p < .001
Evidence	F(1.00, 15104.05) = 0.04, p = .848
Incentive:Group	F(4.00, 112.10) = 2.28, p = .065
Accuracy:Evidence	F(1.00, 15097.33) = 185.32, p < .001
Group:Accuracy	F(2.00, 15106.28) = 2.27, p = .103
Group:Evidence	F(2.00, 15099.41) = 3.51, p = .030
Group:Accuracy:Evidence.	F(2.00, 15094.35) = 3.05, p = .047
Model 4	Confidence
Incentive	F(2.00, 112.34) = 20.93, p < .001
Group	F(2.00, 112.50) = 2.75, p = .068
Sex	F(1.00, 110.26) = 2.88, p = .093
IQ	F(1.00, 109.80) = 0.03, p = .865
Accuracy	F(1.00, 15107.01) = 609.14, p < .001
Evidence	F(1.00, 15104.51) = 0.04, p = .845
Incentive:Group	F(4.00, 112.11) = 2.29, p = .064
Accuracy:Evidence	F(1.00, 15097.16) = 185.42, p < .001
Group:Accuracy	F(2.00, 15106.06) = 2.30, p = .100
Group:Evidence	F(2.00, 15098.91) = 3.45, p = .032
Group:Accuracy:Evidence	F(2.00, 15094.15) = 3.09, p = .046

Table 3: Results of linear mixed-effects models. Shown here are the results of Model 1 (without demographics) and Model 4 (with demographics) acquired using Type 3 F tests with Satterthwaite approximation for degrees of freedom using the afex package. Shown are F values, with corresponding degrees of freedom and P-values.

Model 2: Accuracy ~ Incentive*Group + (1+Incentive|Subject)	
Group	$F_{2,109}$ = 0.5827, P = 0.5601
Incentive	$F_{2,1591}$ = 1.0319, P = 0.3566
Group*Incentive	$F_{4,1586}$ = 0.8671, P = 0.4830
Model 3: RT ~ Incentive*Group + (1+Incentive|Subject)	
Group	$F_{2,110}$ = 0.5207, P = 0.5956
Incentive	$F_{2,220}$ = 0.0994, P = 0.9054
Group*Incentive	$F_{4,219}$ = 0.4269, P = 0.7891

Table 4 | Results of control models. Shown here are the results of Model 2 and Model 3 linear mixed-effects models, acquired using Type 3 F tests with Satterthwaite approximation for degrees of freedom using the afex package. Shown are F values, with corresponding degrees of freedom and P-values

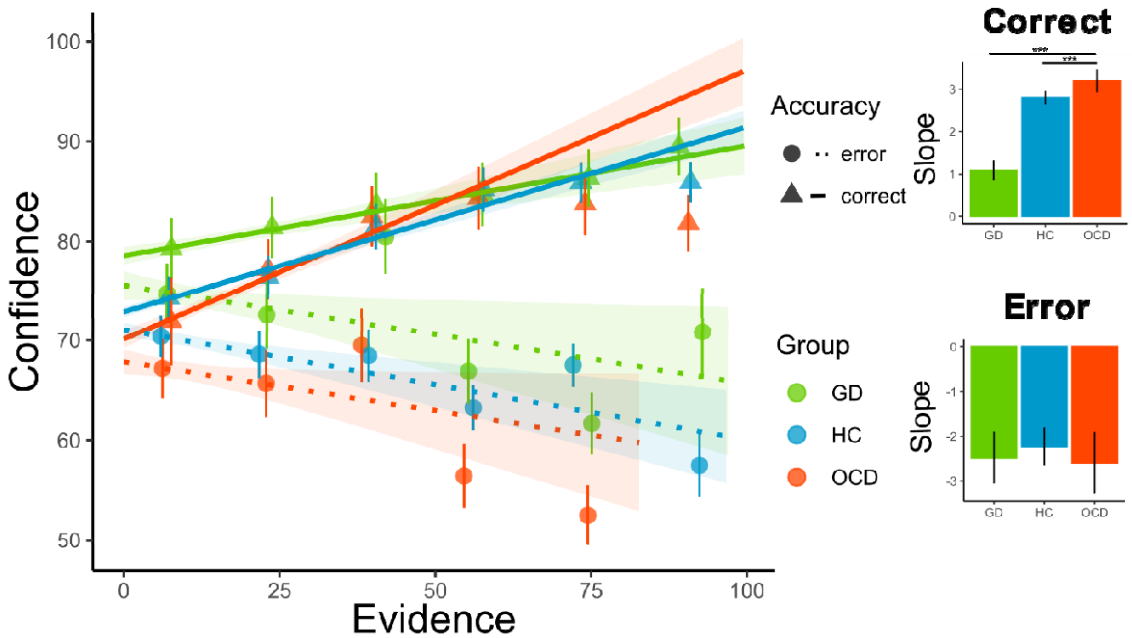

Figure 3 | Evidence integration per group and accuracy level in confidence. Linking evidence, accuracy and group. Triangles represent mean reported confidence as a function of evidence for correct answers, and dots for incorrect answers, with different colors for the three groups. The solid lines represent the best linear regression fit for each group separately at the population level for correct answers, and the dotted lines for incorrect answers. Error bars represent SEM per group, shaded areas represent 95% confidence interval. Insets represent slopes (estimated marginal means of trends, taken from emtrends() function, error bars represent SEM) of correct and incorrect answers per group. Results from post-hoc testing are shown, where the slope for correct answer is significantly lower for gambling disorder (GD) versus both healthy controls (HC) and obsessive-compulsive disorder (OCD) (* p<0.05, ** p<0.01, *** p<0.001).

Since we found that sex and IQ were significantly different between the groups, we aimed to control for these variables by adding them as fixed effects, resulting in Model 4. The results indicated that the main effect of group did not remain significant, but showed a trend towards an effect ($F_{2,112} = 2.7465$, $p=0.06846$), while the main effect of incentive did remain significant ($F_{2,112} = 20.9326$, $p = < 0.001$). We did not find evidence for a significant effect of sex ($F_{1,110} = 2.8776$, $p =0.09264$), or IQ ($F_{1,109} = 0.0291$, $p =0.86489$). The interaction effect between group and incentive remained non-significant at trend-level ($F_{4,112} = 2.2898$, $p = 0.06412$). The significant three-way interaction between accuracy, evidence and group persisted ($F_{2,15094} = 3.0871$, $p=0.04566$). Importantly, when performing a Chi-square test on the log-likelihood values of the models excluding and including the demographic variables to compare model fit, we found that the model without demographics showed a better model

fit (X2 = 2.7018, df=2, p=0.259), thereby favoring this simpler model. Additionally, to investigate how confidence was differently affected by sex in our healthy controls, we performed a two-sample t-test which showed that males were generally more confident than females (males: 76.51 +- 1.04; females: 71.70 +- 0.77) (t_{52} = 2.6518, p-value = 0.01057). However, both sex and IQ did not show a significant influence on confidence level in Model 4.

Between groups fMRI results GLM 1

We analyzed the functional neuroimaging data to test for differences in brain activity between groups for our contrasts of interest: (1) choice moment modulated by early certainty, (2) rating/incentive moment modulated by incentive value, and (3) rating/incentive moment modulated by confidence. The results from the fMRI group analysis revealed no significant differences between the groups in neuronal activity for any of our contrasts.

Across groups fMRI results GLM 1

Next, we grouped all subjects together and performed one-sample t-test on our contrasts of interest to examine the results across groups (cluster-generating voxel threshold P<=.001 uncorr.; clusterwise correction for multiple comparisons P_{FWE}<0.05). During choice, early certainty positively correlated with activation in the precuneus, ventromedial prefrontal cortex, bilateral ventral striatum and putamen, and bilateral visual areas (**Figure 4A**). Areas negatively related to early certainty were the dorsal anterior cingulate cortex, bilateral dorsomedial- and dorsolateral prefrontal cortex, bilateral insula, thalamus, middle frontal gyrus, bilateral sensorimotor cortex, superior and inferior parietal lobe (**Figure 4A**).

At the moment of incentive presentation, the incentive value correlated positively with activation in the ventral striatum and vmPFC stretching into more dorsal areas, as well as the

superior temporal gyrus (**Figure 4B**). Incentive value was negatively related to BOLD activity in the right (pre)motor cortex and dorsolateral PFC, as well as the left middle and superior temporal gyrus, left occipitotemporal gyrus, and left middle and inferior frontal gyrus. Moreover, activity in right lateral occipitotemporal gyrus and middle temporal gyrus were negatively related to incentive value (**Figure 4B**).

During rating moment, confidence was positively related to activity in the vmPFC, left motor cortex and putamen and bilateral visual areas (**Figure 4C**). The following areas were negatively related to confidence: the left superior and inferior parietal lobes, right dorsolateral PFC, right supramarginal gyrus and thalamus, right motor cortex stretching into the dorsolateral PFC, left visual cortex and cerebellum (**Figure 4C**). See **Table 5** for details of across group fMRI results.

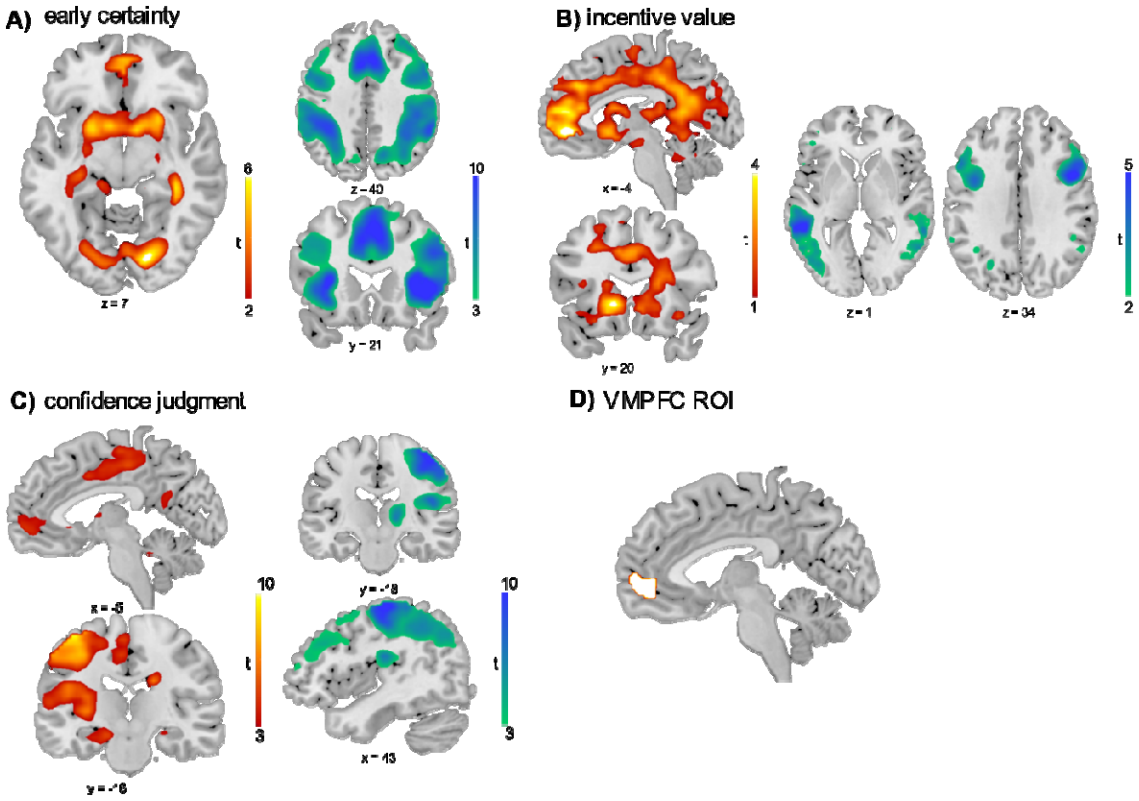

Figure 4 | Whole brain statistical bold-oxygen-level-dependent (BOLD) activity across groups. Red/yellow areas represent areas with a positive relationship, while green/blue areas represent areas

that have a negative relationship. (A) Areas correlating significantly with early certainty at choice moment. Shown are positive activations in ventromedial prefrontal cortex, ventral striatum and visual cortices. Negative activations in dorsal anterior cingulate cortex, dorsolateral prefrontal cortices, insula, parietal cortices. (B) Areas correlating significantly with incentive value at incentive/rating moment. Shown are positive activations in ventromedial prefrontal cortex, anterior cingulate cortex, ventral striatum. Negative activations in dorsolateral prefrontal cortices and temporal gyri (C) Areas correlating significantly with confidence judgments at incentive/rating moment. Positive actions are shown in ventromedial prefrontal cortex, motor cortex and putamen. Negative clusters in motor cortex and dorsolateral prefrontal cortex. All clusters survived P<0.05 FWE cluster correction. Voxel-wise cluster-defining threshold was set at P<.001, uncorrected. For whole brain activation table see table 5. (D) Region of interest (ROI) of the VMPFC used for GLM2 analyses.

Effect	**Brain Region**	**k**	**Peak z-score**	**P (FWE cluster corrected)**	**Peak MNI x**	**y**	**z**	**Hemi-sphere**
Early Certainty +	Precuneus Ventromedial PFC Ventral Striatum Putamen	2180	6.66	<.001	-6	-34	11	LR
	Lingual gyrus (visual cortex)	154	6.39	<.001	18	-81	-4	R
	Lingual gyrus (visual cortex)	54	4.49	0.045	-21	-79	-4	L
Early Certainty -	Dorsal Anterior Cingulate Dorsomedial PFC Dorsolateral PFC Insula Thalamus Middle Frontal Gyrus Precentral Gyrus Postcentral Gyrus Supramarginal Gyrus Superior Parietal Lobe Inferior Parietal Lobe Calcarine gyrus (visual cortex)	13299	Inf (>8)	<.001	45	14	2	LR
	Middle Occipital Lobe	451	7.06	<.001	-30	-91	-4	L

	Middle Temporal Gyrus Lateral Occipito-temporal Gyrus				-48 -45	-67 -61	-1 -10	
	Right Cerebellum	144	6.64	<.001	33	-55	-31	R
Incentive Value +	Ventral Striatum	74	4.75	.004	-12	11	-4	L
	Ventromedial PFC Dorsomedial PFC	212	4.53	<.001	-3 -9 0	44 50 35	-4 -4 14	LR
	Superior Temporal Gyrus	48	4.25	.026	-45 -39	-16 -22	-1 5	L
Incentive Value -	Precentral gyrus stretching into premotor cortex and dorsolateral PFC	283	5.81	<.001	39 45 48	11 5 14	26 32 29	R
	Middle temporal gyrus Superior temporal gyrus	277	5.26	<.001	-54 -51 -48	-43 -52 -25	2 11 -7	L
	Lateral occipitotemporal gyrus Medial occipitotemporal gyrus	183	5.06	<.001	-45 -24 -24	-61 -73 -82	-13 -7 -10	L
	Middle frontal gyrus Inferior frontal gyrus	299	4.93	<.001	-45 -39 -54	2 17 17	53 23 14	L
	Lateral occipitotemporal gyrus	116	4.90	<.001	42 45	-58 -49	-13 -13	R
	Middle temporal gyrus	47	3.74	.029	57 60 57	-46 -46 -61	11 2 2	R
Confidence +	Middle occipitotemporal gyrus Lateral occipitotemporal gyrus Cerebellum	1947	Inf (>8)	<.001	12 21 15	-73 -70 -52	-10 -7 -16	R
	Motor cortex (precentral gyrus)	993	Inf (>8)	<.001	-36	-25	65	L

					-36 -54	-19 -16	47 47	
	Putamen Rolandic operculum	968	5.91	<.001	-30 -45 -30	-19 -16 -22	2 20 14	L
	Occipital lobe	65	4.58	.011	42	-67	5	R
	Ventromedial PFC	92	4.39	.002	-3 -12 -19	56 47 41	-4 8 -1	LR
Confidence -	Lingual gyrus (visual cortex) Cerebellum	1144	Inf (>8)	<.001	-9 -15 -24	-79 -52 -67	-7 -22 -28	L
	Motor cortex (precentral gyrus) Stretching into dorsolateral PFC	2421	Inf (>8)	<.001	45 42 39	-16 -37 -52	59 62 41	R
	Supramarginal gyrus Thalamus	262	6.92	<.001	45 15	-19 -22	20 2	R
	Superior parietal lobe Inferior parietal lobe	168	5.09	<.001	-33 -39 -39	-58 -52 -43	41 47 41	L
	Middle frontal gyrus (Dorsolateral PFC)	71	4.49	.007	-45 -45	32 23	32 35	R

Table 5 | Whole brain activation tables. Brain activations (whole brain analyses) showing activity related to early certainty at choice moment, as well as activity related to incentive and confidence at incentive/rating moment. All whole-brain activation maps were thresholded using family-wise error correction for multiple correction (FWE) at cluster level (P FWE_clu < 0.05), with a voxel cluster-defining threshold of P<0.001 uncorrected. Activity that positively correlates to given variable is denoted by '+', whereas negative correlations are denoted by '-'. PFC = prefrontal cortex.

Interaction between metacognition and motivation in VMPFC (GLM 2)

Our recent study revealed an interaction between motivational and metacognitive signals in the VMPFC, suggesting an important role of the VMPFC in the interaction between metacognition and motivation (Hoven et al., 2020). To investigate how this interaction takes effect in and differs between our clinical groups, we performed an ROI analysis by leveraging our factorial design. We extracted VMPFC activations for both time points (choice and rating), all incentives (loss, neutral and gain), all groups (HC, OCD and GD), for both baseline activity and a regression slope with (1) signed evidence and (2) confidence judgments for all these events (see **Figure 5D** for the ROI).

First, one-sample t-tests showed that, overall, VMPFC baseline activations were negative at choice and rating moment (choice: $t_{100} = -3.611$, $p < 0.001$; baseline: $t_{100} = -4.9287$, $p < 0.001$). The correlations between VMPFC activity and both signed evidence at choice moment and confidence at rating moment, however, were significantly positive (choice: $t_{100} = 3.057$, $p = 0.003$; baseline: $t_{100} = 3.7399$, $p < 0.001$) (**Figure 5**). This implies that the VMPFC represents both confidence judgments and signed evidence (i.e. the interaction between accuracy and evidence, showing that VMPFC activity was lowest when one had high levels of evidence but was incorrect, and highest when one had a lot of evidence and was in fact correct). Then, we turned to see whether there were effects of incentive condition and group around this general signal. As expected, at choice moment there were no effects of incentive condition on VMPFC baseline activity, nor on its correlation with the signed evidence signal (i.e. slope) (**Figure 5, Table 6**). Despite the behavioral group effect on evidence integration, we did not find a group nor an interaction effect on both baseline VMPFC activity and the correlation with signed evidence. At rating moment, however, incentive condition had a significant effect on both the baseline VMPFC activity, as well as its correlation with confidence. Post-hoc testing showed that the baseline VMPFC activity was higher during gain versus loss (t_{196}: -3.874, $p < 0.001$), and during gain versus neutral (t_{196}: -3.228, $p < 0.001$), but no differences between neutral and loss conditions were found (t_{196}: -0.646, p =

0.7948). The correlation of VMPFC activity with confidence was significantly higher (i.e. increased slope) in gain versus neutral (t_{196} = -3.053, p = 0.0072), while no differences between gain and loss, or between neutral and loss were found. Moreover, there was a significant group effect on VMPFC baseline activity during rating moment. The post-hoc tests revealed that OCD subjects had significantly decreased activity compared with HCs, averaged over incentive conditions (t_{98} = -2.515, p = 0.0358). No interaction effects between group and incentive were found on baseline activity or its correlation with confidence at rating moment.

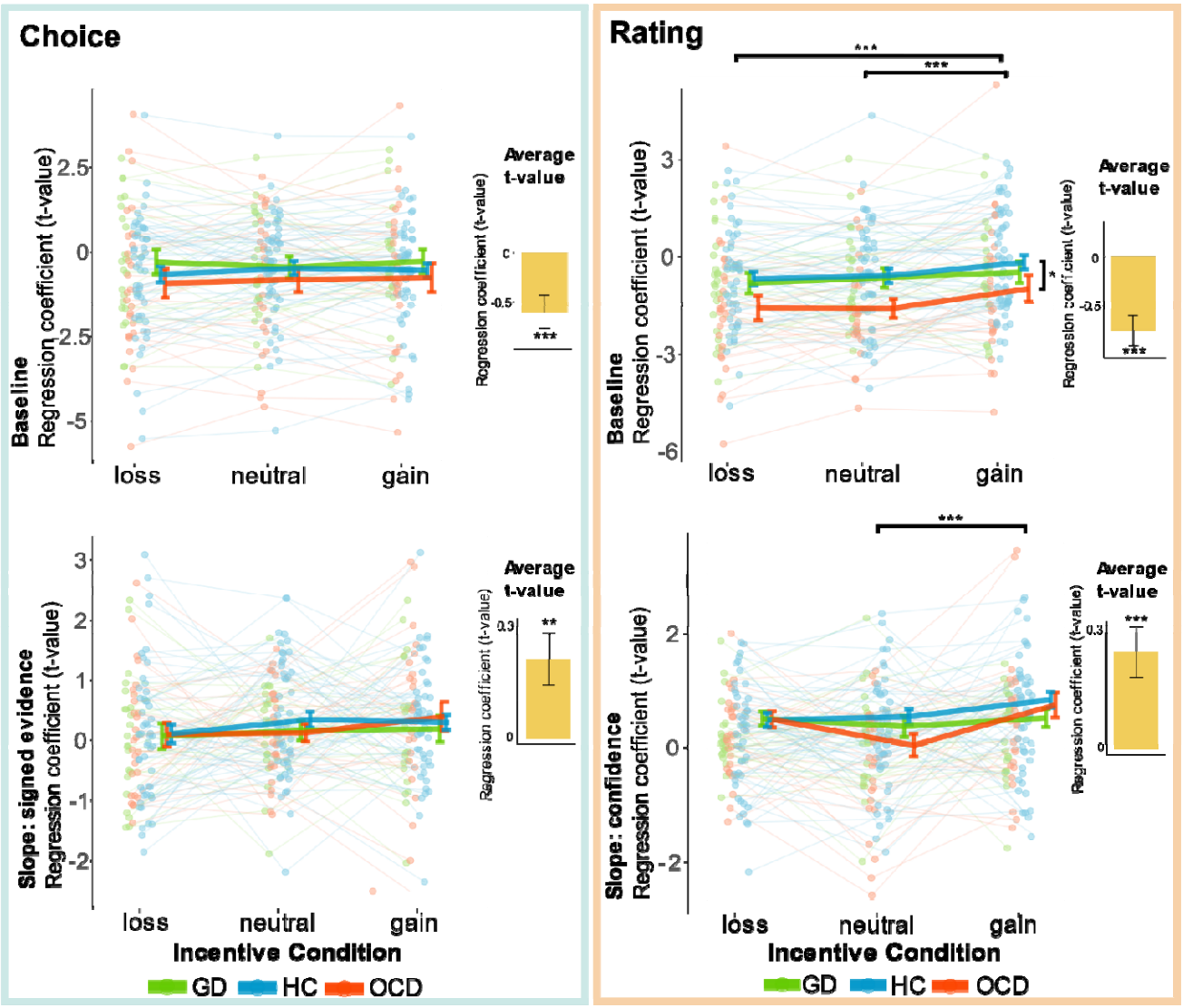

Figure 5 | Ventromedial prefrontal cortex region of interest (ROI) analysis. T-values corresponding to baseline and regression slopes were extracted for all three groups and three incentive conditions, at two time points of interest: choice and incentive/rating moment. Green dots and lines represent gambling disorder patients, blue dots and lines represent healthy controls and red dots and lines represent obsessive-compulsive disorder patients. Dots represent individual t-statistics,

and error bars represent sample mean ± SEM per group. Black bars represent significant post-hoc tests. Yellow bars represent average t-values, with corresponding significance level of one-sample t-tests against 0. (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). GD = gambling disorder, HC = healthy control, OCD = obsessive-compulsive disorder.

	Incentive	**Group**	**Incentive:Group**
Choice Baseline	$F(1.99, 195.28) = 0.37$, $p = 0.687$	$F(2, 98) = 0.54$, $p = 0.582$	$F(3.99, 195.28) = 0.41$, $p = 0.803$
Choice Slope 'Signed Evidence'	$F(1.99, 195) = 1.15$, $p = 0.320$	$F(2, 98) = 0.20$, $p=0.819$	$F(3.98, 195) = 0.31$, $p = 0.869$
Rating Baseline	$F(1.91, 186.81) = 8.61$, $p<$**0.001**	$F(2, 98) = 3.24$, $p =$ **0.044**	$F(3.81, 186.81) = 0.44$, $p = 0.771$
Rating Slope 'Confidence Judgment'	$F(1.92, 187.68) = 4.67$, $p =$ **0.012**	$F(2, 98) = 0.99$, $p = 0.375$	$F(3.83, 187.68) = 1.29$, $p = 0.277$

Table 6: Results of VMPFC ROI analysis. Shown here are the results of the mixed ANOVAs of t-statistics in the ventromedial prefrontal cortex (VMPFC) region of interest (ROI) using the afex package. Shown are the main effects of incentive condition, group and their interaction effect on the choice and rating time points, focusing on both the baseline activity as well as the slope of signed evidence and confidence judgments, respectively. F-values, with corresponding degrees of freedom and p-values are reported.

Discussion

In this study we investigated the role of (neural signatures of) metacognitive ability and its interaction with incentive motivation in two compulsive disorders: OCD and GD. We tested group differences in confidence, the influence of incentive motivation on confidence and the integration of evidence in the confidence signal. Lastly, we also investigated the (group differences in the) neural signature of confidence estimations.

First, we replicated the biasing effect of incentives on confidence estimation in all groups, showing that confidence was higher in the gain context and lower in the loss context, while controlling for incentive differences in performance and reaction times. This is a robust effect, that has now been independently replicated multiple times, both in perceptual decision-making as well as reinforcement learning (Hoven et al., 2020; Lebreton et al., 2018; Lebreton, Bacily, et al., 2019; Ting et al., 2020). Regarding our initial hypotheses on confidence differences between groups and the role of the interaction between motivation and confidence on OCD and GD symptoms, we initially found evidence for a significantly higher confidence in GD patients versus OCD patients, although this effect did not reach our pre-defined significance threshold anymore after controlling for sex and IQ differences between groups. In conclusion, we only found moderate evidence for our hypothesis of group differences in confidence levels, as well as for our hypothesis that incentive motivation would affect confidence judgments differently in the groups. Future research should address the role of the demographic confounding factors more specifically. We also investigated the computational signatures of confidence formation in more detail, and replicated the typical X-shaped relationship between accuracy and evidence. Interestingly, GD patients showed less integration of evidence into their confidence judgments for correct choices compared to both HCs and OCD patients. This suggests that GD patients were less able to use evidence they receive to form confidence judgment about their choices. This decreased sensitivity to objective perceptual evidence could fit GD's symptomatology of cognitive inflexibility (Perandrés-Gómez et al., 2021; van Timmeren et al., 2018), and cognitive distortions

(Ledgerwood et al., 2020; Mallorquí-Bagué et al., 2019). Illusion of control leads pathological gamblers to believe they can predict outcomes, rendering them less influenced by objective evidence, which may promote continuation of (overconfident) gambling behavior (Cowley et al., 2015; Goodie & Fortune, 2013).

Our initial hypothesis about GD and OCD patients was that they would show confidence judgments on the opposite side of the spectrum, that is, increased confidence for GD and decreased confidence for OCD. While we did not find any evidence that the patients groups differed from the HC group, we initially did find that GD patients were more confident than OCD patients, which was along the lines of our expectations. However, this effect seemed to be partly driven by sex and IQ differences between groups, since including these measures as covariates resulted in a trend significant effect. Demographically, the GD group consisted mostly of men, whereas the OCD group showed a more balanced representation of men and women, albeit slightly more skewed to the inclusion of women. This mirrors the prevalence distribution of these disorders, GD is more prevalent amongst men (Howe et al., 2019; Subramaniam et al., 2015; Welte et al., 2017), and being male is considered a strong predictor for the development of gambling problems (Dowling et al., 2017), whereas adult OCD patient populations are more equally prevalent across sexes, with a slight preponderance of women (Swedo et al., 1989). Mirroring our findings of increased confidence in HC male subjects, recent studies have shown that men are more confident in their choices, despite equal performance (Ariel et al., 2018; Rivers et al., 2021). Therefore, the effect of sex might have explained some variance in our data that was originally attributed to the effect of group, but does not fully explain the group differences, as we do find a trend toward a group effect. The importance of taking into account sex and gender as factors in both neuroscience and psychiatry research is increasingly recognized and acted upon (Cahill, 2006). It is now well known that sex differences play a role in the incidence, as well as the treatment and manifestation of clinical symptoms in psychiatric disorders (Cosgrove et al., 2007; Gobinath et al., 2017). Given these findings, it is clinically relevant to

study confidence in psychiatry. However, the precise role of sex and gender in abnormal (meta)cognitive behavior in psychiatric disorders deserves more attention and should be characterized further in future research.

Our data did not show convincing evidence for our initial hypothesis on the interaction between incentives and group, that is, an exaggerated decrease/increase in confidence during loss/gain anticipation in OCD/GD, respectively. However, the group*incentive interaction approached significance, and exploratory analyses did show increased confidence in GD patients compared to both OCD patients and HCs, specifically in the gain condition. This finding agrees with literature demonstrating increased reward sensitivity in GD (Navas et al., 2017; Van Holst et al., 2012). Our incentivized perceptual decision-making task might be better tailored to investigate abnormalities in confidence in the GD compared to the OCD population, since it involves the potential of earning money which may be symptom provoking for GD patients. Confidence in OCD patients has been mostly studied using metamemory paradigms, and abnormalities were found to arise especially in specific contexts such as OCD-relevant situations or in case of heightened feeling of responsibility (Boschen & Vuksanovic, 2007; Bucarelli & Purdon, 2016; Hermans et al., 2008; Moritz et al., 2007; Radomsky et al., 2001; Tolin et al., 2001). Task paradigms characterized as more (disorder) neutral found conflicting evidence: some studies showed decreased confidence in OCD patients compared to HCs, whereas others did not (Hoven et al., 2019). Earlier studies probing confidence in GD are sparse, and whilst they all did show an effect of overconfidence in (sub)clinical problem gamblers, none of the studies actively controlled for performance differences, making it difficult to draw strong conclusions about confidence biases (Brevers et al., 2014; Goodie, 2005; Lakey et al., 2007).

Since this study did not find evidence for confidence differences between either clinical group contrasted with the healthy population, we cannot technically speak of confidence ‘abnormalities’ in these two populations. The clinical groups do however, manifest on opposite sides of the spectrum, as was expected. Future work is necessary to study the link

between compulsivity and confidence more directly. One interesting avenue of research is using a transdiagnostic framework to study metacognition in psychiatry. Indeed, studies have shown that a transdiagnostic factor dimension of ‘compulsive behavior and intrusive thoughts’ (CIT) was positively related to confidence, whereas the ‘anxious-depression’ dimension showed a negative relationship in a healthy online sample (Rouault, Seow, et al., 2018). This is complemented by a recent study showing a decoupling of confidence and behavior and diminished utilizing of perceptual evidence, due to inflated confidence levels in individuals scoring high on CIT (Seow & Gillan, 2020). This latter result is in line with our findings of diminished evidence integration into confidence judgments in GD patients. These transdiagnostic findings seem to be inconsistent with earlier work comparing patients to HCs, showing decreased confidence in OCD. Importantly, comorbidities of anxiety and/or depression are commonplace amongst OCD patients, for which earlier studies mostly did not control, possibly clouding the effects of compulsive behavior versus anxious symptomatology on findings of confidence (Gillan et al., 2017). It must be noted that our patient groups did not have comorbid diagnoses. However, this does not eliminate the possibility that our patients were also suffering from non-diagnosed anxious beliefs alongside compulsive behaviors.

The brain areas we found to be related to confidence and incentive processing converge with earlier work. Confidence was found to be positively related to the vmPFC via automatic processing at the choice moment (De Martino et al., 2013; Lebreton et al., 2015; Lopez-Persem et al., 2020; Shapiro & Grafton, 2020). Early certainty processing was also positively related to activity in the ventral striatum and precuneus (Hebart et al., 2016; Rouault, Mcwilliams, et al., 2018; Vaccaro & Fleming, 2018). We also observed a wide-spread network of areas negatively related to early certainty, containing the dACC, dorsolateral PFC, insula, inferior parietal lobe and midfrontal gyrus, a network repeatedly associated with uncertainty and metacognitive processes (Hebart et al., 2016; Molenberghs et al., 2016; Morales et al., 2018; Vaccaro & Fleming, 2018). Well-known relationships between reward processing and activity in both ventral striatum and vmPFC (Bartra et al., 2013; Lebreton et

al., 2009) were replicated. Moreover, we found negative relationships between incentive value and BOLD activity in the central executive network (i.e. lateral PFC and middle frontal gyrus), as well as superior temporal gyrus (Liu et al., 2011; Wilson et al., 2018). Confidence was found to be related to vmPFC activity, not only at choice moment, but also during rating (De Martino et al., 2013; Lebreton et al., 2015; Lopez-Persem et al., 2020), whereas our earlier fMRI study showed vmPFC activity only during choice moment, with sub-threshold confidence encoding during rating in HCs (Hoven et al., 2020). This is mostly likely due to power issues: our current fMRI sample consists of 104 participants, compared to 30 participants in our earlier work. Overall, our fMRI findings closely resemble activation patterns previously shown in healthy populations.

We also replicated our earlier findings of an effect of incentive condition on VMPFC baseline activity and on the correlation of VMPFC activity with confidence, which was highest in gain conditions (Hoven et al., 2020). Whilst we did find that GD patients showed aberrant evidence integration on a behavioral level, we did not find any group differences in evidence processing on neurobiological level. Interestingly, OCD patients showed a decreased baseline VMPFC activity during incentive/rating moment, which fits with earlier work showing neurobiological deficits in a ‘ventral motivational circuit’ including the VMPFC (Stein et al., 2019; Thorsen et al., 2018). However, we did not find any interactions with incentive condition in the VMPFC activity related to either signed evidence or confidence.

In sum, contrary to our hypotheses, we did not find neurobiological deficits directly related to confidence or to the effects of incentive on confidence in our clinical samples. This might not be surprising, given that the behavioral group effects were small (and disappeared when controlling for demographics), which limited our ability a priori to find impairments in neural circuits mediating confidence processes. Because, to our knowledge, the present study represents the first attempt in investigating the joint neural basis of metacognitive and motivational processes in both GD and OCD, further study - e.g. looking into transdiagnostic variations of symptoms - might be more powerful in detecting clinically useful neurocognitive

signatures of those processes than the present clinical case-control comparisons. For example, a recent study investigating compulsivity with resting-state effective connectivity in the cortical-striatal-thalamic-cortical (CSTC) circuit, found that high levels of compulsivity correlated with disturbances in the CSTC network and that the variance in connectivity was better explained by transdiagnostic variation within the phenotype than within the traditional diagnostic groups of GD and OCD patients, since no effects were found for any of the case-control contrasts (Parkes et al., 2019).

In conclusion, this study replicated the previously demonstrated value-induced confidence bias, and showed some tentative evidence for increased confidence in GD compared to OCD patients, which must be interpreted with caution since controlling for sex diminished this effect. Importantly, since the two patient groups did not significantly differ from the healthy population, we did not find evidence for abnormal behavioral confidence processes, nor for divergent influence of incentive value on confidence, nor for differences in the underlying neural processes.

Literature

Admon, R., Bleich-Cohen, M., Weizmant, R., Poyurovsky, M., Faragian, S., & Hendler, T. (2012). Functional and structural neural indices of risk aversion in obsessive-compulsive disorder (OCD). *Psychiatry Research - Neuroimaging*, *203*(2–3), 207–213. https://doi.org/10.1016/j.pscychresns.2012.02.002

Allen, M., Frank, D., Samuel Schwarzkopf, D., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *ELife*, *5*, 1–17. https://doi.org/10.7554/eLife.18103

Allen, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., Callaghan, M. F., & Rees, G. (2017). Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage*, *149*, 415–423. https://doi.org/10.1016/j.neuroimage.2017.02.008

Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there sex differences in confidence and metacognitive monitoring accuracy for everyday, academic, and psychometrically measured spatial ability? *Intelligence*, *70*, 42–51. https://doi.org/10.1016/j.intell.2018.08.001

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *Journal of Neuroscience*, *33*(42), 16657–16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6082–6087. https://doi.org/10.1073/pnas.1800795115

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bey, K., Lennertz, L., Riesel, A., Klawohn, J., Kaufmann, C., Heinzel, S., Grützmann, R., Kathmann, N., & Wagner, M. (2017). Harm avoidance and childhood adversities in patients with obsessive-compulsive disorder and their unaffected first-degree relatives. *Acta Psychiatrica Scandinavica*, *135*(4), 328–338. https://doi.org/10.1111/acps.12707

Bey, K., Weinhold, L., Grützmann, R., Heinzel, S., Kaufmann, C., Klawohn, J., Riesel, A., Lennertz, L., Schmid, M., Ramirez, A., Kathmann, N., & Wagner, M. (2020). The polygenic risk for obsessive-compulsive disorder is associated with the personality trait harm avoidance. *Acta Psychiatrica Scandinavica*, *142*(4), 326–336. https://doi.org/10.1111/acps.13226

Black, D. W., & Shaw, M. (2019). The Epidemiology of Gambling Disorder. *Gambling Disorder*, 29–48. https://doi.org/10.1007/978-3-030-03060-5_3

Boschen, M. J., & Vuksanovic, D. (2007). Deteriorating memory confidence, responsibility perceptions and repeated checking: Comparisons in OCD and control samples. *Behaviour Research and Therapy*, *45*(9), 2098–2109. https://doi.org/10.1016/j.brat.2007.03.009

Brevers, D., Cleeremans, A., Bechara, A., Greisen, M., Kornreich, C., Verbanck, P., & Noël, X. (2014). Impaired Metacognitive Capacities in Individuals with Problem Gambling. *Journal of Gambling Studies*, *30*(1), 141–152. https://doi.org/10.1007/s10899-012-9348-3

Bucarelli, B., & Purdon, C. (2016). Stove checking behaviour in people with OCD vs. anxious controls.

Journal of Behavior Therapy and Experimental Psychiatry, *53*, 17–24. https://doi.org/10.1016/j.jbtep.2016.03.005

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*(6), 477–484. https://doi.org/10.1038/nrn1909

Calado, F., & Griffiths, M. D. (2016). Problem gambling worldwide: An update and systematic review of empirical research (2000-2015). *Journal of Behavioral Addictions*, *5*(4), 592–613. https://doi.org/10.1556/2006.5.2016.073

Chamberlain, S. R., Blackwell, A. D., Fineberg, N. A., Robbins, T. W., & Sahakian, B. J. (2005). The neuropsychology of obsessive compulsive disorder: The importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neuroscience and Biobehavioral Reviews*, *29*(3), 399–419. https://doi.org/10.1016/j.neubiorev.2004.11.006

Choi, J. S., Shin, Y. C., Jung, W. H., Jang, J. H., Kang, D. H., Choi, C. H., Choi, S. W., Lee, J. Y., Hwang, J. Y., & Kwon, J. S. (2012). Altered Brain Activity during Reward Anticipation in Pathological Gambling and Obsessive-Compulsive Disorder. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045938

Clark, L., Boileau, I., & Zack, M. (2019). Neuroimaging of reward mechanisms in Gambling disorder: an integrative review. *Molecular Psychiatry*, *24*(5), 674–693. https://doi.org/10.1038/s41380-018-0230-2

Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry. *Biological Psychiatry*, *62*(8), 847–855. https://doi.org/10.1016/j.biopsych.2007.03.001

Cowley, E., Briley, D. A., & Farrell, C. (2015). How do gamblers maintain an illusion of control? *Journal of Business Research*, *68*(10), 2181–2188. https://doi.org/10.1016/J.JBUSRES.2015.03.018

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. https://doi.org/10.1038/nn.3279

Dowling, N. A., Merkouris, S. S., Greenwood, C. J., Oldenhof, E., Toumbourou, J. W., & Youssef, G. J. (2017). Early risk and protective factors for problem gambling: A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, *51*, 109–124. https://doi.org/10.1016/j.cpr.2016.10.008

Ferris, J., & Wynne, H. (2001). The Canadian Problem Gambling Index: Final report. *Canadian Centre on Substance Abuse*, 38.

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., Potenza, M. N., Robbins, T. W., & Denys, D. (2016). Compulsivity in obsessive-compulsive disorder and addictions. *European Neuropsychopharmacology*, *26*(5), 856–868. https://doi.org/10.1016/j.euroneuro.2015.12.003

Figee, M., Vink, M., De Geus, F., Vulink, N., Veltman, D. J., Westenberg, H., & Denys, D. (2011). Dysfunctional reward circuitry in obsessive-compulsive disorder. *Biological Psychiatry*, *69*(9), 867–874. https://doi.org/10.1016/j.biopsych.2010.12.003

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286.

https://doi.org/10.1098/rstb.2012.0021

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883.Relating

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2017). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1). https://doi.org/10.1038/s41562-016-0002

Fortune, E. E., & Goodie, A. S. (2012). Cognitive distortions as a component and treatment focus of pathological gambling: A review. *Psychology of Addictive Behaviors*, *26*(2), 298–310. https://doi.org/10.1037/a0026422

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*, *106*, 134–143. https://doi.org/10.1016/j.neuroimage.2014.11.036

Gherman, S., & Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *ELife*, *7*. https://doi.org/10.7554/eLife.38293

Gillan, C. M., Fineberg, N. A., & Robbins, T. W. (2017). A trans-diagnostic perspective on obsessive-compulsive disorder. *Psychological Medicine*, *47*(9), 1528–1548. https://doi.org/10.1017/S0033291716002786

Gobinath, A. R., Choleris, E., & Galea, L. A. M. (2017). Sex, hormones, and genotype interact to influence psychiatric disease, treatment, and behavioral research. *Journal of Neuroscience Research*, *95*(1–2), 50–64. https://doi.org/10.1002/jnr.23872

Goodie, A. S. (2005). The role of perceived control and overconfidence in pathological gambling. *Journal of Gambling Studies*, *21*(4), 481–502. https://doi.org/10.1007/s10899-005-5559-1

Goodie, A. S., & Fortune, E. E. (2013). Measuring cognitive distortions in pathological gambling: Review and meta-analyses. *Psychology of Addictive Behaviors*, *27*(3), 730–743. https://doi.org/10.1037/a0031892

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: I. Development, Use, and Reliability. *Archives of General Psychiatry*, *46*(11), 1006–1011. https://doi.org/10.1001/archpsyc.1989.01810110048007

Grant, J. E., Odlaug, B. L., & Chamberlain, S. R. (2016). Neural and psychological underpinnings of gambling disorder: A review. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *65*, 188–193. https://doi.org/10.1016/j.pnpbp.2015.10.007

Haber, S. N., & Behrens, T. E. J. (2014). The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. *Neuron*, *83*(5), 1019–1039. https://doi.org/10.1016/J.NEURON.2014.08.031

Haber, S. N., & Knutson, B. (2009). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology 2010 35:1*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Hamilton, M. (1959). The Assessment of Anxiety States by Rating. *British Journal of Medical Psychology*, *32*(1), 50–55. https://doi.org/10.1111/j.2044-8341.1959.tb00467.x

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2016). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. *Cerebral Cortex*, *26*(1), 118–130. https://doi.org/10.1093/cercor/bhu181

Hermans, D., Engelen, U., Grouwels, L., Joos, E., Lemmens, J., & Pieters, G. (2008). Cognitive confidence in obsessive-compulsive disorder: Distrusting perception, attention and memory. *Behaviour Research and Therapy*, *46*(1), 98–113. https://doi.org/10.1016/j.brat.2007.11.001

Hilgenstock, R., Weiss, T., & Witte, O. W. (2014). You'd Better Think Twice: Post-Decision Perceptual Confidence. *NeuroImage*, *99*, 323–331. https://doi.org/10.1016/j.neuroimage.2014.05.049

Hoven, M., Brunner, G., De Boer, N. S., Goudriaan, A., Denys, D., Van Holst, R. J., Luigjes, J., & Lebreton, M. (2020). How motivational signals disrupt metacognitive signals in the human VMPFC. *BioRxiv*, 2020.10.02.323550. https://doi.org/10.1101/2020.10.02.323550

Hoven, M., Lebreton, M., Engelmann, J. B., Denys, D., Luigjes, J., & van Holst, R. J. (2019). Abnormalities of confidence in psychiatry: an overview and future perspectives. *Translational Psychiatry*, *9*(1), 1–18. https://doi.org/10.1038/s41398-019-0602-7

Howe, P. D. L., Vargas-Sáenz, A., Hulbert, C. A., & Boldero, J. M. (2019). Predictors of gambling and problem gambling in Victoria, Australia. *PLOS ONE*, *14*(1), e0209277. https://doi.org/10.1371/journal.pone.0209277

Jung, W. H., Kang, D. H., Han, J. Y., Jang, J. H., Gu, B. M., Choi, J. S., Jung, M. H., Choi, C. H., & Kwon, J. S. (2011). Aberrant ventral striatal responses during incentive processing in unmedicated patients with obsessive-compulsive disorder. *Acta Psychiatrica Scandinavica*, *123*(5), 376–386. https://doi.org/10.1111/j.1600-0447.2010.01659.x

Kaufmann, C., Beucke, J. C., Preuße, F., Endrass, T., Schlagenhauf, F., Heinz, A., Juckel, G., & Kathmann, N. (2013). Medial prefrontal brain activation to anticipated reward and loss in obsessive-compulsive disorder. *NeuroImage: Clinical*, *2*(1), 212–220. https://doi.org/10.1016/j.nicl.2013.01.005

Koellinger, P., & Treffers, T. (2015). Joy leads to overconfidence, and a simple countermeasure. *PLoS ONE*, *10*(12), 1–22. https://doi.org/10.1371/journal.pone.0143263

Lakey, C. E., Goodie, A. S., & Campbell, W. K. (2007). Frequent card playing and pathological gambling: The utility of the Georgia Gambling Task and Iowa Gambling Task for predicting pathology. *Journal of Gambling Studies*, *23*(3), 285–297. https://doi.org/10.1007/s10899-006-9034-4

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167. https://doi.org/10.1038/nn.4064

Lebreton, M., Bacily, K., Palminteri, S., & Engelmann, J. B. (2019). Contextual influence on confidence judgments in human reinforcement learning. *PLoS Computational Biology*, *15*(4), e1006973. https://doi.org/10.1371/journal.pcbi.1006973

Lebreton, M., Bavard, S., Daunizeau, J., & Palminteri, S. (2019). Assessing inter-individual differences with task-related functional neuroimaging. *Nature Human Behaviour*, *3*(9), 897–905. https://doi.org/10.1038/s41562-019-0681-8

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation

System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Lebreton, M., Langdon, S., Slieker, M. J., Nooitgedacht, J. S., Goudriaan, A. E., Denys, D., van Holst, R. J., & Luigjes, J. (2018). Two sides of the same coin: Monetary incentives concurrently improve and bias confidence judgments. *Science Advances*, *4*(5), eaaq0668. https://doi.org/10.1126/sciadv.aaq0668

Ledgerwood, D. M., Dyshniku, F., McCarthy, J. E., Ostojic-Aitkens, D., Forfitt, J., & Rumble, S. C. (2020). Gambling-Related Cognitive Distortions in Residential Treatment for Gambling Disorder. *Journal of Gambling Studies*, *36*(2), 669–683. https://doi.org/10.1007/s10899-019-09895-4

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package*, *3*. https://doi.org/10.1080/00031305.1980.10483031

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., & Pessiglione, M. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nature Neuroscience*, *23*(5), 664–675. https://doi.org/10.1038/s41593-020-0615-9

Luigjes, J., Lorenzetti, V., de Haan, S., Youssef, G. J., Murawski, C., Sjoerds, Z., van den Brink, W., Denys, D., Fontenelle, L. F., & Yücel, M. (2019). Defining Compulsive Behavior. *Neuropsychology Review*, *29*(1), 4–13. https://doi.org/10.1007/s11065-019-09404-9

Mallorquí-Bagué, N., Vintró-Alcaraz, C., Verdejo-García, A., Granero, R., Fernández-Aranda, F., Magaña, P., Mena-Moreno, T., Aymamí, N., Gómez-Peña, M., Del Pino-Gutiérrez, A., Mestre-Bach, G., Menchón, J. M., & Jiménez-Murcia, S. (2019). Impulsivity and cognitive distortions in different clinical phenotypes of gambling disorder: Profiles and longitudinal prediction of treatment outcomes. *European Psychiatry*, *61*, 9–16. https://doi.org/10.1016/j.eurpsy.2019.06.006

Massoni, S. (2014). Emotion as a boost to metacognition: How worry enhances the quality of confidence. *Consciousness and Cognition*, *29*, 189–198. https://doi.org/10.1016/j.concog.2014.08.006

Mathes, B. M., Morabito, D. M., & Schmidt, N. B. (2019). Epidemiological and Clinical Gender Differences in OCD. *Current Psychiatry Reports 2019 21:5*, *21*(5), 1–7. https://doi.org/10.1007/S11920-019-1015-2

Mazaika, P., Hoeft, F., Glover, G.H., Reiss, A.L. (2009). Methods and Software for fMRI Analysis for Clinical Subjects. Human Brain Mapping Conference.

Meng, Y. J., Deng, W., Wang, H. Y., Guo, W. J., Li, T., Lam, C., & Lin, X. (2014). Reward pathway dysfunction in gambling disorder: A meta-analysis of functional magnetic resonance imaging studies. *Behavioural Brain Research*, *275*, 243–251. https://doi.org/10.1016/j.bbr.2014.08.057

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Perspective Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*, 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Molenberghs, P., Trautwein, F.-M., Böckler, A., Singer, T., & Kanske, P. (2016). Neural correlates of metacognitive ability and of feeling confident: a large-scale fMRI study. *Social Cognitive and*

Affective Neuroscience, *11*(12), 1942–1951. https://doi.org/10.1093/scan/nsw093

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*, *38*(14), 2360–17. https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Moritz, S., Wahl, K., Zurowski, B., Jelinek, L., Hand, I., & Fricke, S. (2007). Enhanced perceived responsibility decreases metamemory but not memory accuracy in obsessive-compulsive disorder (OCD). *Behaviour Research and Therapy*, *45*(9), 2044–2052. https://doi.org/10.1016/j.brat.2007.03.003

Navas, J. F., Billieux, J., Perandrés-Gómez, A., López-Torrecillas, F., Cándido, A., & Perales, J. C. (2017). Impulsivity traits and gambling cognitions associated with gambling preferences and clinical status. *International Gambling Studies*, *17*(1), 102–124. https://doi.org/10.1080/14459795.2016.1275739

Nestadt, G., Kamath, V., Maher, B. S., Krasnow, J., Nestadt, P., Wang, Y., Bakker, A., & Samuels, J. (2016). Doubt and the decision-making process in obsessive-compulsive disorder. *Medical Hypotheses*, *96*, 1–4. https://doi.org/10.1016/j.mehy.2016.09.010

Parkes, L., Tiego, J., Aquino, K., Braganza, L., Chamberlain, S. R., Fontenelle, L. F., Harrison, B. J., Lorenzetti, V., Paton, B., Razi, A., Fornito, A., & Yücel, M. (2019). Transdiagnostic variations in impulsivity and compulsivity in obsessive-compulsive disorder and gambling disorder correlate with effective connectivity in cortical-striatal-thalamic-cortical circuits. *NeuroImage*, *202*. https://doi.org/10.1016/j.neuroimage.2019.116070

Perandrés-Gómez, A., Navas, J. F., van Timmeren, T., & Perales, J. C. (2021). Decision-making (in)flexibility in gambling disorder. *Addictive Behaviors*, *112*, 106534. https://doi.org/10.1016/j.addbeh.2020.106534

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235. https://doi.org/10.1002/mrm.20900

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. https://doi.org/10.1038/nn.4240

Radomsky, A. S., Rachman, S., & Hammond, D. (2001). Memory bias, confidence and responsibility in compulsive checking. *Behaviour Research and Therapy*, *39*(7), 813–822. https://doi.org/10.1016/S0005-7967(00)00079-6

Rivers, M. L., Fitzsimmons, C. J., Fisk, S. R., Dunlosky, J., & Thompson, C. A. (2021). Gender differences in confidence during number-line estimation. *Metacognition and Learning*, *16*(1), 157–178. https://doi.org/10.1007/s11409-020-09243-7

Romanczuk-Seiferth, N., Koehler, S., Dreesen, C., Wüstenberg, T., & Heinz, A. (2015). Pathological gambling and alcohol dependence: Neural disturbances in reward and loss avoidance processing. *Addiction Biology*, *20*(3), 557–569. https://doi.org/10.1111/adb.12144

Rouault, M., & Fleming, S. M. (2020). Formation of global self-beliefs in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(44), 27268–27276. https://doi.org/10.1073/pnas.2003094117

Rouault, M., Mcwilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, *1*,

17. https://doi.org/10.1017/pen.2018.16

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry*, *84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2008). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry 2010 15:1*, *15*(1), 53–63. https://doi.org/10.1038/mp.2008.94

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *19*(4), 25–25. https://doi.org/10.1167/19.4.25

Samuels, J., Bienvenu, O. J., Krasnow, J., Wang, Y., Grados, M. A., Cullen, B., Goes, F. S., Maher, B., Greenberg, B. D., McLaughlin, N. C., Rasmussen, S. A., Fyer, A. J., Knowles, J. A., Nestadt, P., McCracken, J. T., Piacentini, J., Geller, D., Pauls, D. L., Stewart, S. E., … Nestadt, G. (2017). An investigation of doubt in obsessive–compulsive disorder. *Comprehensive Psychiatry*, *75*, 117–124. https://doi.org/10.1016/j.comppsych.2017.03.004

Sanders, J. I., Hangya, B. B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*, *90*(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Seow, T. X. F., & Gillan, C. M. (2020). Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-59646-4

Sescousse, G., Barbalat, G., Domenech, P., & Dreher, J. C. (2013). Imbalance in the sensitivity to different types of rewards in pathological gambling. *Brain*, *136*(8), 2527–2538. https://doi.org/10.1093/brain/awt126

Shapiro, A. D., & Grafton, S. T. (2020). Subjective value then confidence in human ventromedial prefrontal cortex. *PLOS ONE*, *15*(2), e0225617. https://doi.org/10.1371/journal.pone.0225617

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(20), 22–33. https://doi.org/10.1016/S0924-9338(99)80239-9

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1286–1291. https://doi.org/10.1038/nn.4384

Singmann, H., Bolker, B., & Westfall, J. (2015). *Analysis of Factorial Experiments, package "afex."* 1–44. http://cran.r-project.org/package=afex

Stein, D. J. (2002). Obsessive-compulsive disorder. *Lancet*, *360*(9330), 397–405. https://doi.org/10.1016/S0140-6736(02)09620-4

Stein, D. J., Costa, D. L. C., Lochner, C., Miguel, E. C., Reddy, Y. C. J., Shavitt, R. G., van den Heuvel, O. A., & Simpson, H. B. (2019). Obsessive–compulsive disorder. *Nature Reviews Disease Primers 2019 5:1*, *5*(1), 1–21. https://doi.org/10.1038/s41572-019-0102-3

Subramaniam, M., Wang, P., Soh, P., Vaingankar, J. A., Chong, S. A., Browning, C. J., & Thomas, S. A. (2015). Prevalence and determinants of gambling disorder among older adults: A systematic review. *Addictive Behaviors*, *41*, 199–209. https://doi.org/10.1016/j.addbeh.2014.10.007

Summerfeldt, L. J., Kloosterman, P. H., Antony, M. M., & Swinson, R. P. (2014). Examining an obsessive-compulsive core dimensions model: Structural validity of harm avoidance and

incompleteness. *Journal of Obsessive-Compulsive and Related Disorders*, *3*(2), 83–94. https://doi.org/10.1016/j.jocrd.2014.01.003

Swedo, S. E., Rapoport, J. L., Leonard, H., Lenane, M., & Cheslow, D. (1989). Obsessive-Compulsive Disorder in Children and Adolescents: Clinical Phenomenology of 70 Consecutive Cases. *Archives of General Psychiatry*, *46*(4), 335–341. https://doi.org/10.1001/archpsyc.1989.01810040041007

Thorsen, A. L., Hagland, P., Radua, J., Mataix-Cols, D., Kvale, G., Hansen, B., & van den Heuvel, O. A. (2018). Emotional Processing in Obsessive-Compulsive Disorder: A Systematic Review and Meta-analysis of 25 Functional Neuroimaging Studies. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(6), 563–571. https://doi.org/10.1016/j.bpsc.2018.01.009

Ting, C. C., Palminteri, S., Engelmann, J. B., & Lebreton, M. (2020). Robust valence-induced biases on motor response and confidence in human reinforcement learning. *Cognitive, Affective and Behavioral Neuroscience*, *20*(6), 1184–1199. https://doi.org/10.3758/s13415-020-00826-0

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., Amir, N., Street, G. P., & Foa, E. B. (2001). Memory and memory confidence in obsessive – compulsive disorder. *Behaviour Research and Therapy*, *39*, 913–927. https://doi.org/10.1016/S0005-7967(00)00064-4

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 239821281881059. https://doi.org/10.1177/2398212818810591

Van Holst, R. J., Veltman, D. J., Bchel, C., Van Den Brink, W., & Goudriaan, A. E. (2012). Distorted expectancy coding in problem gambling: Is the addictive in the anticipation? *Biological Psychiatry*, *71*(8), 741–748. https://doi.org/10.1016/j.biopsych.2011.12.030

van Timmeren, T., Daams, J. G., van Holst, R. J., & Goudriaan, A. E. (2018). Compulsivity-related neurocognitive performance deficits in gambling disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *84*, 204–217. https://doi.org/10.1016/j.neubiorev.2017.11.022

Welte, J. W., Barnes, G. M., Tidwell, M. C. O., & Wieczorek, W. F. (2017). Predictors of Problem Gambling in the U.S. *Journal of Gambling Studies*, *33*(2), 327–342. https://doi.org/10.1007/s10899-016-9639-1

Wilson, R. P., Colizzi, M., Bossong, M. G., Allen, P., Kempton, M., Abe, N., Barros-Loscertales, A. R., Bayer, J., Beck, A., Bjork, J., Boecker, R., Bustamante, J. C., Choi, J. S., Delmonte, S., Dillon, D., Figee, M., Garavan, H., Hagele, C., Hermans, E. J., … Bhattacharyya, S. (2018). The Neural Substrate of Reward Anticipation in Health: A Meta-Analysis of fMRI Findings in the Monetary Incentive Delay Task. *Neuropsychology Review*, *28*(4), 496–506. https://doi.org/10.1007/s11065-018-9385-5