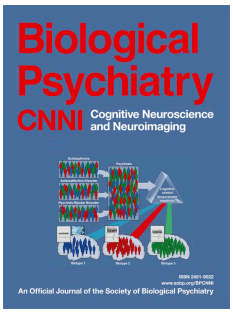
Journal Pre-proof



Neural correlates of metacognition impairment in opioid addiction

Scott J. Moeller, Sameera Abeykoon, Pari Dhayagude, Benjamin Varnas, Jodi J. Weinstein, Greg Perlman, Roberto Gil, Stephen M. Fleming, Anissa Abi-Dargham

PII: S2451-9022(24)00202-7

DOI: https://doi.org/10.1016/j.bpsc.2024.07.014

Reference: BPSC 1263

To appear in: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*

Received Date: 6 June 2024

Revised Date: 15 July 2024

Accepted Date: 19 July 2024

Please cite this article as: Moeller S.J., Abeykoon S., Dhayagude P., Varnas B., Weinstein J.J., Perlman G., Gil R., Fleming S.M. & Abi-Dargham A., Neural correlates of metacognition impairment in opioid addiction, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2024), doi: https://doi.org/10.1016/j.bpsc.2024.07.014.

Title: Neural correlates of metacognition impairment in opioid addiction

Abbreviated Title: Metacognition and opioid addiction

Authors: Scott J. Moeller[1*a], Sameera Abeykoon[1], Pari Dhayagude[2], Benjamin Varnas[1], Jodi J. Weinstein[1], Greg Perlman[1], Roberto Gil[1], Stephen M. Fleming[3,4,5], Anissa Abi-Dargham[1]

Author Affiliations: [1]Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, 11794; [2]University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, WC1B 5EH, United Kingdom; [4]Wellcome Centre for Human Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, United Kingdom; [5]Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, United Kingdom
*Correspondence to: Scott J. Moeller, Health Sciences Center, Level 10, Room 087H, 101 Nicolls Road, Stony Brook, NY 11794-8101, Phone: 631-638-3223, scott.moeller@stonybrookmedicine.edu
[a] First and senior author

Manuscript Information:
Text: 3996 words
Abstract: 245 words
Display Items: 2 Tables, 5 Figures
Supplement: 1228 words, 1 Table, 2 Figures

Key Words: fMRI; metacognition; cognitive neuroscience; clinical neuroscience; opioid use disorder; addiction

ABSTRACT

Background: Individuals with substance use disorder show impaired self-awareness of ongoing behavior. This deficit suggests problems with metacognition, operationalized in the cognitive neuroscience literature as the ability to monitor and evaluate the success of one's own cognition and behavior. However, the neural mechanisms of metacognition have not been characterized in a drug-addicted population. *Methods*: Community samples of participants with opioid use disorder (OUD) (N=27) and healthy controls (N=29) performed a previously-validated fMRI metacognition task (perceptual decision-making task along with confidence ratings of performance). Measures of recent drug use and addiction severity were also acquired. *Results*: Individuals with OUD had lower metacognitive sensitivity than controls (i.e., disconnection between task performance and task-related confidence). Trial-by-trial analyses showed that this overall group difference was driven by (suboptimally) low confidence in OUD during correct trials. In fMRI analyses, the task engaged an expected network of brain regions (e.g., rostrolateral prefrontal cortex and dorsal anterior cingulate/supplementary motor area, both previously linked to metacognition); group differences emerged in a large ventral anterior cluster that included the medial and lateral orbitofrontal cortex and striatum (higher activation in OUD). Trial-by-trial fMRI analyses showed group differences in rostrolateral prefrontal cortex activation, which further correlated with metacognitive behavior across all participants. Exploratory analyses suggested that the behavioral and neural group differences were exacerbated by recent illicit opioid use and unexplained by general cognition. *Conclusions*: With confirmation and extension of these findings,

metacognition and its associated neural circuits could become new, promising therapeutic targets in addiction.

INTRODUCTION

Substance use disorder (SUD) involves taking drugs in larger amounts and for longer periods of time than intended (1), suggesting compromised self-awareness of ongoing behavior. In support, drug-addicted individuals exhibit problems in self-monitoring and self-reporting their ongoing task behavior (2-14) and emotional experiences (15-17). They also underestimate the extent of their cognitive problems when compared with the reports of close informants (18, 19) or other independent observers (20), and they report ambivalence about needing drug treatment despite documented impairment (21-23).

An integrative framework for characterizing deficits of this kind is metacognition, referring to the ability to monitor and evaluate one's own cognition and behavior (e.g., successfully discriminate mistakes from successes). Laboratory studies of metacognition have been mainly conducted in healthy controls (HC). In one representative paradigm, participants trial-by-trial perform a simple cognitive process (e.g., two-choice perceptual decision) and then render a confidence judgment about their performance (24, 25). Metacognition is then operationalized as the degree to which higher confidence correlates with better performance (26-29). Translational evidence from human and preclinical studies has convincingly linked metacognition to functional and structural integrity of the anterior prefrontal cortex (PFC), perhaps especially the rostrolateral PFC (rlPFC) (24, 25, 30-41).

We suggest that metacognition and its neural mechanisms are impaired in opioid use disorder (OUD). We previously reported metacognitive deficits in actively-using individuals with cocaine use disorder, which were further correlated with lower gray

matter volume in the anterior PFC (rostral anterior cingulate cortex) (9). Since then, a different research group reported a similar metacognitive deficit in methadone-maintained individuals with OUD (42). Importantly, deficits in metacognition-related functions, which are referred to in the literature as 'Type 2' cognition, are separable from deficits in more classically-examined cognitive functions [e.g., working memory, sustained attention, or decision-making (43-49)], which are referred to in the literature as 'Type 1' cognition; Type 1 and Type 2 cognition have distinct psychological and neural mechanisms in HC (26, 28, 50-53). Given this separability between Type 1 and Type 2 cognition, we posit that metacognition could serve as a new therapeutic target in drug addiction. However, the functional neural circuitry of metacognition impairment has not been characterized in a drug-addicted population.

Here, to test for impaired metacognition and associated neural abnormalities in OUD, we used an fMRI metacognition task that members of our team previously validated in HC, wherein better metacognition was correlated with stronger confidence-related functional signals in the rlPFC (54). We hypothesized that, compared with HC, OUD participants would show (A) worse metacognition and (B) aberrant confidence-related signals in rlPFC. Although we focused *a priori* on the rlPFC as our main region of interest (ROI), we also tested whole-brain effects. Finally, building on our prior findings in cocaine use disorder (9), we hypothesized that (C) the behavioral and neural measurements would correlate with recency of illicit opioid use in OUD participants.

METHODS

Participants

We acquired data in 30 individuals with OUD and 30 HC, recruited through advertisements, local treatment facilities, and word-of-mouth; all provided written informed consent. Three OUD participants and one HC participant were excluded for excessive motion during fMRI scanning [framewise displacement (FD) (55) ≥ 1.2mm in ≥10% of the task volumes during a given run, and ≥1.2mm in ≥5% of all task volumes], resulting in a final sample of 27 OUD and 29 HC (Table 1). Due to demographic mismatching on race between OUD and HC groups after motion-related exclusions, we controlled for race in all analyses.

Inclusion criteria for all participants were: (A) males and females ages 18-55; (B) English-speaking, for task and questionnaire completion; and (C) good current medical and psychiatric health based on a medical physical and routine blood work as determined by the study psychiatrist. An additional inclusion criterion for OUD participants was (D) history of OUD, with ongoing treatment so that participants were sufficiently stable for brain imaging (treatments included buprenorphine: *n*=18; methadone: *n*=6; naltrexone/other: *n*=3). Exclusion criteria were: (A) head trauma with loss of consciousness >10 minutes; (B) clinically significant medical, neurological, or psychiatric illness that would compromise safety, study completion, or data quality, other than additional SUD in the OUD group and/or nicotine use disorder in either group; (C) medication use within 6 months that would alter cerebral function or otherwise adversely affect the imaging data, except for those used to treat OUD and its sequelae (note that we aimed to recruit a highly generalizable, community sample of OUD who are often on

multiple medications); (D) positive urine toxicology for non-prescribed drugs of abuse except for opioids or cannabis (see Table 1 for urine toxicology results); (E) staff impressions of acute intoxication/impairment; (F) contraindications to MRI; and (G) pregnancy (urine verified) or breast-feeding in women.

All participants underwent a clinical interview, which confirmed OUD diagnosis in the patient sample (with 14 meeting past-year criteria) and lack of psychiatric diagnoses in HC [Structured Clinical Interview for DSM-5 (SCID-5), Research Version (56)] (see Supplement for OUD comorbidities). The clinical interview additionally included: NIDA-Modified ASSIST v2.0 and Timeline Follow-back Calendar (57), together characterizing illicit opioid use in the last 3 months (as a measure of recency); and well-validated instruments of opioid severity, including the Addiction Severity Index (ASI) (58), Desires for Drug Questionnaire (59), and Leeds Dependence Scale (60).

fMRI Task Design

During each trial, after a fixation cross, participants made a two-choice discrimination judgment, categorizing a noisy image with varying amounts of overlaid white noise as either a face or a house (Figure 1) [see also (54)]. Each judgment was followed by a confidence rating. In the “Report” condition, participants rated their decision confidence on a 6-point scale. In the “Follow” condition, instead of rating their decision confidence, participants placed the cursor between two vertical lines, specified by the program (active control condition with similar visual and motor demands). There were 4 runs each containing 75 trials (5 sequences each comprised of 10 Report trials and 5 subsequent Follow trials). Task performance was adaptively controlled (staircase), expected to converge at 71% accuracy. The entire sample approached this target

(*M*=74.3%±10.3), though (Type 1) *d'* was unexpectedly lower in OUD than HC [$F(1,50)=14.75$, $p<0.001$] (Figure 2C). Participants responded during 288.0±23.1 trials (96.0%) with no group differences ($p=0.96$), indicating high task engagement.

MRI Acquisition

MRI scanning was performed on a 3T Prisma^{Fit} (Siemens, Erlangen, Germany) using a 64-channel head-and-neck coil. Four runs of multiband (61) BOLD-sensitive echo-planar imaging (EPI) T2*-weighted task imaging were acquired. Each run comprised 1187 volumes, lasting 16.38min. We used: multiband acceleration of 6 and no GRAPPA, 2mm isotropic voxels, 204mm FOV, 66 slices, 60° FA, and TR/TE=800/25ms. A T1w scan was also acquired using a 3D-MPRAGE sequence: TR/TE/TI =2400/2.24/1060ms, FOV=256, voxel size=0.8×0.8×0.8 mm^3, flip angle=8°, slices=208, and GRAPPA parallel imaging factor=2 (see Supplement for more details).

MRI Preprocessing

Data were preprocessed using the HCP (62) Minimal Preprocessing Pipelines v4.2, smoothed with a 6mm FWHM Gaussian kernel (see Supplement for full details). In addition to the participant exclusions described above, we excluded Run 3 from one OUD participant due to a large number of missed trials, and one HC participant only completed two Runs due to scanner-related discomfort.

Data Analysis

Task Behavior

We used a linear mixed model (LMM), with trials nested within participants, to predict confidence levels trial-by-trial, testing whether OUD participants are overconfident or underconfident relative to HC, forming the basis of a metacognitive impairment. The

LMM had the following predictors: Correctness (i.e., whether the Face/House judgment was correct, per trial: Yes, No), Diagnosis (OUD, HC), and the Correctness×Diagnosis interaction. Analyses were restricted to Report trials; Follow trials are invalid for the computation, as they do not incorporate participants' own confidence judgments.

To confirm the trial-by-trial analyses and to obtain person-level metrics for correlational analyses, we also computed two summary metrics of metacognitive accuracy (again, only incorporating the Report trials) (28, 63). Meta-*d'* was fit to each participant's confidence rating data using maximum likelihood estimation (64). Meta-*d'* is a measure of *metacognitive sensitivity* (i.e., how much participants can discriminate their own correct from incorrect judgments), expressed in the same units as Type 1 sensitivity (*d'*) (i.e., the degree of perceptual accuracy). We also computed *metacognitive efficiency*, defined as meta-*d'*/*d'* ('m-ratio'), which corrects meta-*d'* by task performance. Both metrics, sensitivity and efficiency, were then compared between the groups using GLMs (which controlled for race; dummy coded). All behavioral analyses were considered significant at $p<0.05$.

BOLD-fMRI Analyses

A hemodynamic response function was convolved with a boxcar function spanning the time of the confidence rating. The boxcar was separated into two regressors, one for Report trials and one for Follow trials. The 6 motion parameters (3 rotation, 3 translation) and their temporal first derivatives provided by the HCP preprocessing were included as regressors of no interest.

Next, the Report regressor was parametrically modulated by participants' confidence ratings trial-by-trial, creating the 1st Level contrasts of interest in SPM12.

The 1st Level contrast was then analyzed at the 2nd Level (group level) using an independent t-test in SPM12 while also controlling for race. Using this model, we examined (A) which regions activate in relation to trial-by-trial confidence across all participants, and (B) which regions show differing trial-by-trial activations between OUD and HC.

Finally, we used the Report and Follow regressors to create a 1st Level activation map for the contrast of Report>Follow, a complementary analytical approach. At the 2nd Level, we similarly used an independent t-test in SPM12, controlling for race. This contrast examined (A) which regions activate during confidence ratings in all participants, and (B) which regions show different activations between OUD and HC. To uncover the source of significant Report>Follow effects, posthoc analyses were conducted which modeled the Report and Follow conditions separately.

For both analytical approaches (parametric modulation and Report>Follow), we used both whole-brain voxelwise and ROI approaches. Voxelwise analyses were considered significant using a $p<0.001$ voxelwise-uncorrected threshold and a $p<0.05$ cluster-corrected threshold. The ROI analyses were conducted using unbiased bilateral 8mm spheres of the rlPFC, centered at MNI coordinates x=-33, y=44, z=28 and x=27, y=53, z=25 (Figure 3B), which are the peak coordinates from the trial-by-trial analyses obtained from our prior study (54), thus firmly *a priori*. For these ROI, we conducted one-sample t-tests and GLMs in SPSS, considered significant at $p<0.05$ (uncorrected). All SPM and ROI analyses of between-group differences controlled for race.

Correlation Analyses

First, we tested brain-behavior correlations, examining whether metacognitive sensitivity and (separately) metacognitive efficiency correlated with the rlPFC ROIs and/or (extracted) whole-brain cluster-corrected activations. Second, we tested correlations of select task variables with drug use. Specifically, the behavioral and imaging variables which showed significant differences between the groups in the above analyses were tested here for association with opioid craving, dependence severity, and recent use. The rationale for restricting correlations to those variables first showing between-group differences was: (A) to keep the number of analyses manageable overall; and (B) these would be the behaviors/regions to provide plausible mechanisms of impairment in OUD.

Craving and dependence severity were assessed with the Desires for Drug Questionnaire and the Leeds Dependence Scale, respectively (60), using Pearson correlations in OUD only. Recent illicit opioid use, which was our main interest considering our prior work in cocaine use disorder (9), was ascertained using the NIDA-Modified ASSIST drug use measure which was further cross-checked with a Timeline Follow-back Calendar (and the medical physical if needed for final confirmation), which both ask about drug (opioid) use in the 3 months prior to the study. Due to most OUD reporting no drug use prior to the study, we dichotomized recent illicit opioid use in OUD as follows: any use (OUD+: *N*=7) versus no use (OUD-: *N*=20) in the last 3 months. We used GLMs with linear contrasts to test for graded effects in the behavioral and imaging variables (OUD+, OUD-, HC), controlling for race.

In all analyses, correlations with metacognitive sensitivity or efficiency (primary behavioral metrics) and with left/right rlPFC (primary imaging ROIs) were considered

significant at $p<0.05$ (uncorrected), given our *a priori* hypotheses with these variables. For testing correlations with additional brain activations, we applied a Benjamini-Hochberg correction to control the false discovery rate (FDR) at 5% ($q<0.05$) (65).

RESULTS

Behavior

We conducted a LMM analysis with trial-by-trial confidence as the dependent variable, and with Correctness, Diagnosis, and their interaction as predictors. There was a main effect of Correctness [$\chi^2(1)=868.86$, $p<0.001$; confidence, as to be expected, was higher on correct trials than incorrect trials] but no main effect of Diagnosis [$\chi^2(1)=2.87$, $p=0.09$]. Of greater interest, however, the Correctness×Diagnosis interaction reached significance [$\chi^2(1)=21.11$, $p<0.001$]. Follow-up comparisons showed that OUD participants had lower levels of confidence in their performance than HC on correct trials [$\chi^2(1)=3.87$, $p=0.049$] but not on incorrect trials ($p=0.49$). Furthermore, while all participants showed greater confidence during correct than incorrect trials (i.e., as demonstrated by the Correctness main effect), the difference in confidence levels between correct and incorrect trials was greater for HC [$\chi^2(1)=541.76$, $p<0.001$] than for OUD [$\chi^2(1)=329.12$, $p<0.001$] (Figure 2A). Taken together, OUD participants underestimated their task performance when correct, exhibiting a metacognitive deficit.

This suboptimal confidence in OUD was reflected in the metacognitive sensitivity summary statistic, where OUD participants had lower meta-*d*' than HC, as hypothesized [$F(1,50)=5.67$ $p=0.021$] (Figure 2B). Unexpectedly, the groups did not differ on metacognitive efficiency (meta-*d*'/*d*') [$F(1,50)=0.60$ $p=0.44$] (Figure 2B).

fMRI: Trial-by-Trial Parametric Modulation by Confidence (Table 2)

Across all participants, trial-by-trial confidence ratings were negatively correlated with BOLD-fMRI activity in the left occipital and parietal cortices (Figure 3A, C-D). Exploratory ROI analyses of the extracted activations indicated that the trial-by-trial correlation was numerically lower, albeit not significantly so, in OUD than HC in the inferior occipital [$F(1,50)=3.26$ $p=0.077$] and parietal [$F(1,50)=4.00$ $p=0.051$] cortices. In rlPFC ROI analyses, all participants showed a negative trial-by-trial correlation between confidence and brain activity in both rlPFC ROIs [left: one-sample $t(55)=2.11$, $p=0.039$]; right: one-sample $t(55)=2.23$, $p=0.030$]. Of greater interest, these two rlPFC ROIs differed between the groups, where the correlation between confidence and activity was more negative in OUD than HC [left: $F(1,50)=4.35$ $p=0.042$; right: $F(1,50)=6.02$ $p=0.018$] (Figure 3E).

fMRI: Group Activation Mapping Differences for Report>Follow

The Supplement provides the activations and deactivations to the Report>Follow contrast across all participants. When specifically examining group differences to Report>Follow, OUD had greater activation than HC in a large cluster of primarily left ventral and anterior brain areas, with the strongest signal observed in the left orbitofrontal cortex (OFC) extending into the striatum (Figure 4B). Posthoc analyses suggested that this group difference was mostly driven by less deactivation in HC during the Follow condition. However, this posthoc analysis did not reach cluster-level significance and therefore is not interpreted further. Similarly, the rlPFC ROIs for the contrast Report>Follow did not differ between the groups (both $F<0.17$, $p>0.68$).

Correlation Analyses

Brain-Behavior Correlations

Metacognitive sensitivity was tested for association with the OFC/striatal ROI (from the Report>Follow analyses) and bilateral rlPFC ROIs (from the parametric modulation analyses). Across all participants, metacognitive sensitivity was positively correlated with the right rlPFC signals: the less negative the trial-by-trial correspondence between brain activation and confidence, the greater the behavioral metacognition (r=0.27, p=0.044) (Figure 3E). The correlation was not significant in either OUD or HC groups considered separately (both p>0.21). A follow-up robust regression analysis, conducted given the presence of potential outliers, confirmed the correlation across the sample (b=0.46, SE=0.23, p=0.049). The left rlPFC showed a similar, but nonsignificant trend (r=0.22, p=0.108). Finally, there was a similar, but not FDR-corrected correlation between parietal-confidence signals and metacognition (r=0.30, p=0.023).

Relationships with Recency and Severity of Illicit Opioid Use

For recent drug use, significant linear contrasts (groups: OUD+, OUD-, HC) emerged for metacognitive sensitivity (M_{diff}=0.55, SE=0.17, p=0.002), Report>Follow OFC/striatal activation (M_{diff}=0.62, SE=0.14, p<0.001), left rlPFC parametric modulation (M_{diff}=0.40, SE=0.10, p<0.001), and right rlPFC parametric modulation (M_{diff}=0.39, SE=0.10, p<0.001) (Figure 5). Whereas the OFC/striatum Report>Follow linear effect reflected a case-control difference, the other linear effects reflected modulation by recent drug use. That is, compared with OUD-, OUD+ participants had lower metacognitive sensitivity (p=0.031), more strongly negative left rlPFC confidence signals (p=0.002), and more strongly negative right rlPFC confidence signals (p=0.007).

In contrast, neither severity of dependence (Leeds) nor craving (DDQ) was correlated with metacognitive sensitivity, OFC Report>Follow activation, or left/right rlPFC confidence signals (all $p>0.098$).

DISCUSSION

Using a previously validated fMRI task (54), we hypothesized and found that individuals with OUD had worse metacognitive sensitivity than HC, reflecting a poorer trial-by-trial mapping between task accuracy and confidence in that accuracy. Unexpectedly, the groups only differed on metacognitive sensitivity, not efficiency. That is, when controlling for subtle differences in first-order performance (*d'*) between the groups, the group differences in metacognitive sensitivity were no longer apparent. This result was unexpected because prior addiction studies, including our own, had reported group differences in metacognitive efficiency (9, 42). One key difference between our study and prior studies is the current fMRI component, which necessitated a task design where the decision-making (Type 1) portion of the trial was not self-paced. Slow performance, which was registered as "incorrect" on our fMRI task, could have been one reason for lower task performance in OUD participants. Future studies could incorporate additional practice trials outside the scanner, to better calibrate the staircase. Nevertheless, our study, which uses a laboratory-based task, extends research on "metacognitive beliefs." This parallel body of research uses self-reports to ask people to reflect on whether their thoughts are controllable and affect subsequent behavior (66, 67), investigated in the context of alcohol consumption (68, 69), smoking dependence (70), and problematic use of cannabis (71). Our study, which uses a

laboratory-based task of metacognition reduces concerns about demand characteristics (72) and socially desirable responding (73).

We did not initially anticipate that the metacognition impairment in OUD would be driven by underconfidence. Gambling addiction and intoxicated driving, for instance, have been linked to overconfidence (74, 75). Our findings could reflect a distinctive treatment-seeking OUD phenotype, where for example heroin users show less risk-taking than cocaine users (76), especially among those who initiate medications for opioid use disorder (MOUD) (77). Our current OUD sample also may have been experiencing residual withdrawal, stress reactivity, and dysphoria/negative emotionality [i.e., hyperkatifeia (78)], and many patients have low confidence that they can manage such symptoms (79). This direction of effects is also consistent with data acquired in individuals with anxiety and depression symptoms (but without OUD), who show metacognitive underconfidence (80) that is rescued by cognitive behavioral therapy (81). Future studies will need to confirm underconfidence in OUD while also examining relationships with other laboratory tasks and real-world functioning. For example, underconfidence in OUD may be related to suboptimal information-seeking (82), a form of impaired self-regulation. Underconfidence may also affect quality-of-life, recognized as an important clinical outcome that is complementary to drug abstinence (83). Underconfidence in OUD could relate to outcomes such as greater self-stigma, unwarranted hesitancy to take adaptive risks (e.g., accepting a job promotion, starting a family, etc.), or low self-efficacy in the ability to cope with life challenges or remain in drug treatment (84-87).

In the fMRI data, parametric modulation analyses showed significant (negative) correlations between activation and confidence in the left occipital and parietal cortices across all participants. These findings agree with a recent study showing that negative confidence during task performance tracked with activation in the parietal cortex (88). A subsequent coordinate-based meta-analysis similarly identified the parietal cortex as parametrically correlating with confidence (37). Although our study showed no whole-brain between-group differences in these regions (or others), group effects did emerge in rlPFC ROI analyses. These rlPFC ROI analyses also showed negative correlations (i.e., significantly less than 0) in all participants, consistent with prior work (54, 89, 90), but in this region the activations correlating with confidence were especially negative in OUD, resulting in a case-control difference. Furthermore, the weaker (less negative) the trial-by-trial correspondence between (right) rlPFC activation and confidence, the better was the metacognition in all participants. This is opposite to what we previously observed in HC (54), where better metacognitive sensitivity was linked to stronger (more negative) modulation of rlPFC activity by confidence. The reason for this difference is unclear, but we note here that it was the OUD group, not the HC group, which showed a baseline negative rlPFC BOLD signal in relation to confidence. Furthermore, interpreting mass-univariate differences in PFC activation in relation to metacognition is nuanced by findings that the multivariate voxel pattern in both medial and lateral anterior PFC also tracks confidence (33). Despite these variations in directionality, our collective results suggest that confidence-PFC signals, which are critical for effective metacognition as shown in basic research (24, 25, 30-41), are disrupted in OUD, possibly providing a neural basis for the behavioral impairment.

Additional BOLD-fMRI analyses revealed that confidence judgments activated a diverse network of brain regions across all participants (Supplement), which were consistent with the prior report of this task (54). With respect to differences between OUD and HC on this Report>Follow contrast, we observed a large cluster with peak activation in the left lateral OFC extending to the medial PFC/ventromedial PFC, subgenual ACC, striatum, and parahippocampal gyrus. Emerging research has pointed toward ventral prefrontal and/or striatal regions as having a role in confidence (74, 88-93), self-performance estimates (94), and metacognition (95). One interpretation is that this activation, especially in the more ventromedial PFC portions of the cluster, could be pointing to case-control differences in the neural circuits subserving self-awareness (6, 96), which is needed for intact metacognition.

Finally, a subset of the behavioral and neural effects were modulated by past 3-month illicit opioid use. OUD+ participants were the most impaired on metacognitive sensitivity and most dissimilar from HC in their bilateral rlPFC-confidence signals. Interestingly, this modulation was specific to recency of drug use, not seen for severity. These results highlighting recent use effects on metacognition agree with our prior work in cocaine use disorder (9). Metacognition and related functions are also impaired among acutely-intoxicated users of cannabis (97) and alcohol (98), though in our study participants were not acutely intoxicated. Importantly, recent drug use and metacognition are both modifiable variables; metacognition can be trained, and the effects may generalize to additional tasks and self-regulatory domains (99, 100). It remains an open question whether enhancements in metacognition behavior and associated circuits drive functional improvement in psychopathology (i.e., as part of a

causal model, as we would hypothesize) or whether metacognitive enhancements simply track with an otherwise successful treatment response. Future clinical trials can test metacognitive training as a potential adjunctive treatment for OUD, which could help adjudicate between these competing hypotheses and also potentially have clinical benefit.

This study has several limitations. First, most OUD participants were taking medications (MOUD and others, such as antidepressants), and most OUD participants had drug-related and/or psychiatric comorbidities (Table 1; Supplement). This limitation is balanced, though, by the generalizability of this community sample, reflecting the kinds of patients seen in clinics. Second, task accuracy unexpectedly differed between the groups. Future studies could increase response windows and/or include a pre-scan staircasing phase. Third, future studies need to verify the recent drug use effects, using larger samples with more equally-balanced OUD subgroups. Fourth, this study cross-sectional study cannot speak to whether the metacognition behavioral and neural abnormalities predate or follow drug use. Fifth, we only examined metacognition for perceptual decision-making. Although current evidence indicates that metacognition is at least partially domain-general (33) – that is, all types of metacognition invoke behavioral self-evaluation in the form of ‘propositional confidence’ (29) – future studies in OUD will need to examine other metacognitive domains, such as confidence in one’s memory, subjective value assessments, or action capability. Such efforts will aid clinical characterization and eventual intervention.

In conclusion, we showed metacognitive and associated neural abnormalities in a community-based, generalizable sample OUD. Individuals with OUD, and especially

OUD+, were underconfident in their ongoing task performance, a deficit which was linked with aberrant confidence-related activation patterns in the rlPFC, a region critical for metacognition. If these findings are (A) confirmed with larger samples, (B) verified to be dissociable from Type 1 cognition, (C) predictive of real-world functional outcomes, and (D) modifiable with drug abstinence, then metacognition could ultimately become a promising therapeutic target for OUD.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by grants from the National Institute on Drug Abuse (R01DA051420, R01DA049733, R21DA051179, and R21DA048196 to SJM; R01DA057268 to GP) and the National Institute of Mental Health (K23MH115291 to JJW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests

The authors report no biomedical financial interests or potential conflicts of interest.

Author Contributions

SJM designed the study, with scientific input from SF and AA-D. PD, BV, JJW, GP, and RG conducted study procedures. SJM, SA, and GP analyzed data. SJM wrote the paper. All authors read, edited, and approved the final draft.

References

1. American, Psychiatric, Association (2013): *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: Author.

2. Moeller SJ, Maloney T, Parvaz MA, Alia-Klein N, Woicik PA, Telang F, et al. (2010): Impaired insight in cocaine addiction: laboratory evidence and effects on cocaine-seeking behaviour. *Brain*. 133:1484-1493.

3. Hester R, Simões-Franklin C, Garavan H (2007): Post-error behavior in active cocaine users: Poor awareness of errors in the presence of intact performance adjustments. *Neuropsychopharmacology*. 32:1974-1984.

4. Hester R, Nestor L, Garavan H (2009): Impaired error awareness and anterior cingulate cortex hypoactivity in chronic cannabis users. *Neuropsychopharmacology*. 34:2450-2458.

5. Le Berre AP, Muller-Oehring EM, Kwon D, Serventi MR, Pfefferbaum A, Sullivan EV (2016): Differential compromise of prospective and retrospective metamemory monitoring and their dissociable structural brain correlates. *Cortex*. 81:192-202.

6. Moeller SJ, Goldstein RZ (2014): Impaired self-awareness in human addiction: deficient attribution of personal relevance. *Trends Cogn Sci*. 18:635-641.

7. Goldstein RZ, Craig AD, Bechara A, Garavan H, Childress AR, Paulus MP, et al. (2009): The neurocircuitry of impaired insight in drug addiction. *Trends Cogn Sci*. 13:372-380.

8. Le Berre AP, Sullivan EV (2016): Anosognosia for Memory Impairment in Addiction: Insights from Neuroimaging and Neuropsychological Assessment of Metamemory. *Neuropsychol Rev*.

9. Moeller SJ, Fleming SM, Gan G, Zilverstand A, Malaker P, dOleire Uquillas F, et al. (2016): Metacognitive impairment in active cocaine use disorder is associated with individual differences in brain structure. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*. 26:653-662.

10. Moeller SJ, Maloney T, Parvaz MA, Dunning JP, Alia-Klein N, Woicik PA, et al. (2009): Enhanced choice for viewing cocaine pictures in cocaine addiction. *Biol Psychiatry*. 66:169-176.

11. Rupp CI, Derntl B, Osthaus F, Kemmler G, Fleischhacker WW (2017): Impact of Social Cognition on Alcohol Dependence Treatment Outcome: Poorer Facial Emotion Recognition Predicts Relapse/Dropout. *Alcohol Clin Exp Res*. 41:2197-2206.

12. Philippot P, Kornreich C, Blairy S, Baert I, Den Dulk A, Le Bon O, et al. (1999): Alcoholics' deficits in the decoding of emotional facial expression. *Alcohol Clin Exp Res*. 23:1031-1038.

13. Liu Y, Wang L, Yu C, Liu M, Li H, Zhang Y, et al. (2022): How drug cravings affect metacognitive monitoring in methamphetamine abusers. *Addict Behav*. 132:107341.

14. Soutschek A, Bulley A, Wittekind CE (2022): Metacognitive deficits are associated with lower sensitivity to preference reversals in nicotine dependence. *Sci Rep*. 12:19787.

15. Payer DE, Lieberman MD, London ED (2011): Neural correlates of affect processing and aggression in methamphetamine dependence. *Arch Gen Psychiatry*. 68:271-282.

16. Moeller SJ, Konova AB, Parvaz MA, Tomasi D, Lane RD, Fort C, et al. (2014): Functional, structural, and emotional correlates of impaired insight in cocaine addiction. *JAMA Psychiatry*. 71:61-70.

17. Parvaz MA, Moeller SJ, Goldstein RZ (2016): Incubation of Cue-Induced Craving in Adults Addicted to Cocaine Measured by Electroencephalography. *JAMA Psychiatry*. 73:1127-1134.

18. Verdejo-Garcia A, Perez-Garcia M (2008): Substance abusers' self-awareness of the neurobehavioral consequences of addiction. *Psychiatry Res*. 158:172-180.

19. Moreno-Lopez L, Albein-Urios N, Martinez-Gonzalez JM, Soriano-Mas C, Verdejo-Garcia A (2017): Neural correlates of impaired self-awareness of apathy, disinhibition and dysexecutive deficits in cocaine-dependent individuals. *Addict Biol*. 22:1438-1448.

20. Noël X, Saeremans M, Kornreich C, Chatard A, Jaafari N, D'Argembeau A (2022): Reduced calibration between subjective and objective measures of episodic future thinking in alcohol use disorder. *Alcohol Clin Exp Res*. 46:300-311.

21. Moeller SJ, Kundu P, Bachi K, Maloney T, Malaker P, Parvaz MA, et al. (2020): Self-awareness of problematic drug use: Preliminary validation of a new fMRI task to assess underlying neurocircuitry. *Drug Alcohol Depend*. 209:107930.

22. Maremmani AG, Rovai L, Rugani F, Pacini M, Lamanna F, Bacciardi S, et al. (2012): Correlations between awareness of illness (insight) and history of addiction in heroin-addicted patients. *Frontiers in psychiatry / Frontiers Research Foundation*. 3:61.

23. Probst C, Manthey J, Martinez A, Rehm J (2015): Alcohol use disorder severity and reported reasons not to seek treatment: a cross-sectional study in European primary care practices. *Subst Abuse Treat Prev Policy*. 10:32.

24. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010): Relating introspective accuracy to individual differences in brain structure. *Science*. 329:1541-1543.

25. Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014): Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*.

26. Fleming SM, Daw ND (2017): Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychol Rev*. 124:91-114.

27. Fleming SM, Dolan RJ, Frith CD (2012): Metacognition: computation, biology and function. *Philos Trans R Soc Lond B Biol Sci*. 367:1280-1286.

28. Fleming SM, Lau HC (2014): How to measure metacognition. *Frontiers in human neuroscience*. 8:443.

29. Fleming SM (2024): Metacognition and Confidence: A Review and Synthesis. *Annu Rev Psychol*. 75:241-268.

30. Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A (2014): Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*. 84:190-201.

31. Baird B, Cieslak M, Smallwood J, Grafton ST, Schooler JW (2015): Regional white matter variation associated with domain-specific metacognitive accuracy. *J Cogn Neurosci*. 27:440-452.

32. Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010): Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn Neurosci*. 1:165-175.

33. Morales J, Lau H, Fleming SM (2018): Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *J Neurosci*. 38:3534-3546.

34. Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, et al. (2010): Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience research*. 68:199-206.

35. Bang D, Fleming SM (2018): Distinct encoding of decision confidence in human medial prefrontal cortex. *Proc Natl Acad Sci U S A*. 115:6082-6087.

36. Fleming SM, van der Putten EJ, Daw ND (2018): Neural mediators of changes of mind about perceptual decisions. *Nat Neurosci*. 21:617-624.

37. Vaccaro AG, Fleming SM (2018): Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and neuroscience advances*. 2:2398212818810591.

38. Hilgenstock R, Weiss T, Witte OW (2014): You'd better think twice: post-decision perceptual confidence. *Neuroimage*. 99:323-331.

39. McCaig RG, Dixon M, Keramatian K, Liu I, Christoff K (2011): Improved modulation of rostrolateral prefrontal cortex using real-time fMRI training and meta-cognitive awareness. *Neuroimage*. 55:1298-1305.

40. Miyamoto K, Trudel N, Kamermans K, Lim MC, Lazari A, Verhagen L, et al. (2021): Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. *Neuron*. 109:1396-1408.e1397.

41. Yeon J, Shekhar M, Rahnev D (2020): Overlapping and unique neural circuits are activated during perceptual decision making and confidence. *Sci Rep*. 10:20761.

42. Sadeghi S, Ekhtiari H, Bahrami B, Ahmadabadi MN (2017): Metacognitive Deficiency in a Perceptual but Not a Memory Task in Methadone Maintenance Patients. *Sci Rep*. 7:7052.

43. Baldacchino A, Balfour DJ, Passetti F, Humphris G, Matthews K (2012): Neuropsychological consequences of chronic opioid use: a quantitative review and meta-analysis. *Neurosci Biobehav Rev*. 36:2056-2068.

44. Mintzer MZ, Copersino ML, Stitzer ML (2005): Opioid abuse and cognitive performance. *Drug Alcohol Depend*. 78:225-230.

45. Mintzer MZ, Stitzer ML (2002): Cognitive impairment in methadone maintenance patients. *Drug Alcohol Depend*. 67:41-51.

46. Ahn WY, Vasilev G, Lee SH, Busemeyer JR, Kruschke JK, Bechara A, et al. (2014): Decision-making in stimulant and opiate addicts in protracted abstinence: evidence from computational modeling with pure users. *Frontiers in psychology*. 5:849.

47. Maguire DR, Henson C, France CP (2016): Daily morphine administration increases impulsivity in rats responding under a 5-choice serial reaction time task. *British journal of pharmacology*. 173:1350-1362.

48. Myers CE, Sheynin J, Balsdon T, Luzardo A, Beck KD, Hogarth L, et al. (2016): Probabilistic reward- and punishment-based learning in opioid addiction: Experimental and computational data. *Behav Brain Res*. 296:240-248.

49. Ieong HF, Yuan Z (2017): Resting-State Neuroimaging and Neuropsychological Findings in Opioid Use Disorder during Abstinence: A Review. *Frontiers in human neuroscience*. 11:169.

50. Fleming SM, Dolan RJ (2012): The neural basis of metacognitive ability. *Philos Trans R Soc Lond B Biol Sci*. 367:1338-1349.

51. Barrett AB, Dienes Z, Seth AK (2013): Measures of metacognition on signal-detection theoretic models. *Psychological methods*. 18:535-552.

52. Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G (2011): Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious Cogn*. 20:1787-1792.

53. Rouault M, McWilliams A, Allen MG, Fleming SM (2018): Human metacognition across domains: insights from individual differences and neuroimaging. *Personality neuroscience*. 1.

54. Fleming SM, Huijgen J, Dolan RJ (2012): Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci*. 32:6117-6125.

55. Konova AB, Ceceli AO, Horga G, Moeller SJ, Alia-Klein N, Goldstein RZ (2023): Reduced neural encoding of utility prediction errors in cocaine addiction. *Neuron*. 111:4058-4070.e4056.

56. First MB, Williams JBW, Karg RS, Spitzer RL (2015): *Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)*. Arlington, VA: American Psychiatric Association.

57. Miller WR, Del Boca FK (1994): Measurement of drinking behavior using the Form 90 family of instruments. *J Stud Alcohol Suppl*. 12:112-118.

58. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. (1992): The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat*. 9:199-213.

59. Franken IH, Hendriksa VM, van den Brink W (2002): Initial validation of two opiate craving questionnaires the obsessive compulsive drug use scale and the desires for drug questionnaire. *Addict Behav*. 27:675-685.

60. Raistrick D, Bradshaw J, Tober G, Weiner J, Allison J, Healey C (1994): Development of the Leeds Dependence Questionnaire (LDQ): a questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. *Addiction*. 89:563-572.

61. Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, et al. (2010): Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med*. 63:1144-1153.

62. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. (2013): The WU-Minn Human Connectome Project: an overview. *Neuroimage*. 80:62-79.

63. Schulz L, Fleming SM, Dayan P (2023): Metacognitive computations for information search: Confidence in control. *Psychol Rev*. 130:604-639.

64. Maniscalco B, Lau H (2012): A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn*. 21:422-430.

65. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 57:289-300.

66. Hamonniere T, Varescon I (2018): Metacognitive beliefs in addictive behaviours: A systematic review. *Addict Behav*. 85:51-63.

67. Neighbors C, Tomkins MM, Lembo Riggs J, Angosta J, Weinstein AP (2019): Cognitive factors and addiction. *Current opinion in psychology*. 30:128-133.

68. Spada MM, Caselli G, Wells A (2009): Metacognitions as a predictor of drinking status and level of alcohol use following CBT in problem drinkers: a prospective study. *Behav Res Ther*. 47:882-886.

69. Spada MM, Wells A (2009): A metacognitive model of problem drinking. *Clinical psychology & psychotherapy*. 16:383-393.

70. Nikčević AV, Alma L, Marino C, Kolubinski D, Yılmaz-Samancı AE, Caselli G, et al. (2017): Modelling the contribution of negative affect, outcome expectancies and metacognitions to cigarette use and nicotine dependence. *Addict Behav*. 74:82-89.

71. Hamonniere T, Milan L, Varescon I (2022): Repetitive negative thinking, metacognitive beliefs, and their interaction as possible predictors for problematic cannabis use. *Clinical psychology & psychotherapy*. 29:706-717.

72. Williamson A (2007): Using self-report measures in neurobehavioural toxicology: can they be trusted? *Neurotoxicology*. 28:227-234.

73. Crowne DP, Marlowe D (1960): A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*. 24:349-354.

74. Hoven M, de Boer NS, Goudriaan AE, Denys D, Lebreton M, van Holst RJ, et al. (2022): Metacognition and the effect of incentive motivation in two compulsive disorders: Gambling disorder and obsessive-compulsive disorder. *Psychiatry Clin Neurosci*. 76:437-449.

75. Liu L, Chui WH, Deng Y (2021): Driving after alcohol consumption: A qualitative analysis among Chinese male drunk drivers. *Int J Drug Policy*. 90:103058.

76. Bornovalova MA, Daughters SB, Hernandez GD, Richards JB, Lejuez CW (2005): Differences in impulsivity and risk-taking propensity between primary users of crack cocaine and primary users of heroin in a residential substance-use program. *Exp Clin Psychopharmacol*. 13:311-318.

77. Aklin WM, Severtson SG, Umbricht A, Fingerhood M, Bigelow GE, Lejuez CW, et al. (2012): Risk-taking propensity as a predictor of induction onto naltrexone treatment for opioid dependence. *J Clin Psychiatry*. 73:e1056-1061.

78. Koob GF (2022): Anhedonia, Hyperkatifeia, and Negative Reinforcement in Substance Use Disorders. *Current topics in behavioral neurosciences*. 58:147-165.

79. Hall OT, Vilensky M, Teater JE, Bryan C, Rood K, Niedermier J, et al. (2024): Withdrawal catastrophizing scale: initial psychometric properties and implications for the study of opioid use disorder and hyperkatifeia. *Am J Drug Alcohol Abuse*.1-13.

80. Rouault M, Seow T, Gillan CM, Fleming SM (2018): Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biol Psychiatry*. 84:443-451.

81. Fox CA, Lee CT, Hanlon AK, Seow TXF, Lynch K, Harty S, et al. (2023): An observational treatment study of metacognition in anxious-depression. *Elife*. 12.

82. Desender K, Murphy P, Boldt A, Verguts T, Yeung N (2019): A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. *J Neurosci*. 39:3309-3319.

83. Bray JW, Aden B, Eggman AA, Hellerstein L, Wittenberg E, Nosyk B, et al. (2017): Quality of life as an outcome of opioid use disorder treatment: A systematic review. *J Subst Abuse Treat*. 76:88-93.

84. Jones S, Jack B, Kirby J, Wilson TL, Murphy PN (2021): Methadone-Assisted Opiate Withdrawal and Subsequent Heroin Abstinence: The Importance of Psychological Preparedness. *Am J Addict*. 30:11-20.

85. Senbanjo R, Wolff K, Marshall EJ, Strang J (2009): Persistence of heroin use despite methadone treatment: poor coping self-efficacy predicts continued heroin use. *Drug and alcohol review*. 28:608-615.

86. Saffari M, Chang KC, Chen JS, Chang CW, Chen IH, Huang SW, et al. (2022): Temporal associations between depressive features and self-stigma in people with substance use disorders related to heroin, amphetamine, and alcohol use: a cross-lagged analysis. *BMC psychiatry*. 22:815.

87. Bozinoff N, Anderson BJ, Bailey GL, Stein MD (2018): Correlates of Stigma Severity Among Persons Seeking Opioid Detoxification. *J Addict Med*. 12:19-23.

88. Rouault M, Lebreton M, Pessiglione M (2023): A shared brain system forming confidence judgment across cognitive domains. *Cereb Cortex*. 33:1426-1439.

89. Mazor M, Friston KJ, Fleming SM (2020): Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife*. 9.

90. Molenberghs P, Trautwein FM, Bockler 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*. 11:1942-1951.

91. Hoven M, Brunner G, de Boer NS, Goudriaan AE, Denys D, van Holst RJ, et al. (2022): Motivational signals disrupt metacognitive signals in the human ventromedial prefrontal cortex. *Communications biology*. 5:244.

92. De Martino B, Fleming SM, Garrett N, Dolan RJ (2013): Confidence in value-based choice. *Nat Neurosci*. 16:105-110.

93. Lebreton M, Abitbol R, Daunizeau J, Pessiglione M (2015): Automatic integration of confidence in the brain valuation signal. *Nat Neurosci*. 18:1159-1167.

94. Rouault M, Fleming SM (2020): Formation of global self-beliefs in the human brain. *Proc Natl Acad Sci U S A*. 117:27268-27276.

95. Gherman S, Philiastides MG (2018): Human VMPFC encodes early signatures of confidence in perceptual decisions. *Elife*. 7.

96. Maracic CE, Moeller SJ (2021): Neural and Behavioral Correlates of Impaired Insight and Self-Awareness in Substance Use Disorder. *Curr Behav Neurosci Rep*. 8:113-123.

97. Adam KCS, Doss MK, Pabon E, Vogel EK, de Wit H (2020): Δ(9)-Tetrahydrocannabinol (THC) impairs visual working memory performance: a randomized crossover trial. *Neuropsychopharmacology*. 45:1807-1816.

98. Honan CA, Skromanis S, Johnson EG, Palmer MA (2018): Alcohol intoxication impairs recognition of fear and sadness in others and metacognitive awareness of emotion recognition ability. *Emotion*. 18:842-854.

99. Carpenter J, Sherman MT, Kievit RA, Seth AK, Lau H, Fleming SM (2019): Domain-general enhancements of metacognitive ability through adaptive training. *Journal of experimental psychology General*. 148:51-64.

100. Gilbert SJ, Bird A, Carpenter JM, Fleming SM, Sachdeva C, Tsai PC (2020): Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of experimental psychology General*. 149:501-517.

Figure Captions

Figure 1. Task schematic of the fMRI metacognition task. During “Report” trials, participants indicated their confidence trial-by-trial after making a perceptual decision about whether a fuzzy image with varying amounts of white noise is either a face or a house. In the “Follow” condition, all aspects of the trial are identical except that participants do not rate their confidence on a trial, but rather are asked to move the cursor between two blue bars to a location determined by sampling from previous Report trials.

Figure 2. Behavioral results. (A) Trial-by-trial analyses, using a linear mixed model (LMM) with trials nested within participants, showed that individuals with opioid use disorder (OUD) have lower confidence than healthy controls during correct trials, indicating unwarranted pessimism about their performance. (B) This translates into lower metacognitive sensitivity, which reflects a poorer mapping between task accuracy and confidence, though we did not observe group differences in metacognitive efficiency (sensitivity normalized by accuracy). (C) One potential explanation for lack of metacognitive efficiency group differences is that task accuracy (*d'*) also differed between the groups. Asterisks denote $p<0.05$. Estimated marginal means and standard errors are shown.

Figure 3. Trial-by-trial imaging results. Here, activations during the “Report” condition, during which active confidence judgments were made, were parametrically modulated by confidence ratings trial-by-trial. (A, C) Across all participants, there were whole-brain corrected results in the left occipital and parietal cortices, where activations in these regions were negatively correlated with confidence ratings. (B, D) In *a priori* ROI

analyses, the bilateral rostrolateral prefrontal cortex (rlPFC), a well-established region subserving metacognition, also showed negative trial-by-trial correlations with confidence in all participants, and these effects were more pronounced in individuals with opioid use disorder (OUD), resulting in a significant group difference. Estimated marginal means and standard errors are shown. (E) Scatterplot showing that the less negative the trial-by-trial correlation in the rlPFC, the better the metacognition in all participants.

Figure 4. Imaging data for the fMRI contrast of Report>Follow, as modeled during the confidence reporting phase of each trial. (A) Activations across all participants. (B) Group differences in a large (predominantly left) ventral cluster, with peak activation in the left orbitofrontal cortex and extending to the striatum. Estimated marginal means and standard errors are shown.

Figure 5. Modulation by recent illicit opioid use. Individuals with opioid use disorder who used illicit opioids in the past 3 months (Opioid+) had the (A) most impaired metacognition and (B-D) most abnormal (i.e., most different from healthy controls) brain activations in the (B) orbitofrontal/striatal cluster, (C) left rostrolateral prefrontal cortex (rlPFC), and (D) right rlPFC. In all four metrics, individuals with OUD who did not use illicit opioids within months (Opioid-) were intermediate between Opioid+ participants and healthy controls. Estimated marginal means and standard errors are shown.

Table 1. Demographics and clinical characteristics of the study sample at baseline assessment.

Measure	Opioid Use Disorder (N=27)	Healthy Controls (N=29)	Statistical Test
Gender (M / F)	16 / 11	15 / 14	χ^2=0.32
Age (*M* ± *SD*)	34.4 ± 4.8	32.4 ± 9.4	*t*=1.01
Race			χ^2=9.38*
White *N* (%)	24 (88.9)	22 (75.9)	
Black *N* (%)	0 (0.0)	3 (10.3)	
Asian *N* (%)	0 (0.0)	2 (6.9)	
Pacific Islander *N* (%)	0 (0.0)	1 (3.4)	
More than one race *N* (%)	3 (11.1)	1 (3.4)	
Ethnicity (Hispanic / Not Hispanic)	4 / 23	3 / 26	χ^2=0.26
Cigarette and nicotine use			
Any nicotine use *N* (%)	23 (85.2)	8 (27.6)	χ^2=20.17*
Cigarette smoker *N* (%)	18 (66.7)	8 (27.6)	χ^2=8.81*
Cigarettes per day (in smokers)	10.4 ± 8.7	17.4 ± 11.6	*t*=1.71
Patient Health Questionnaire	9.0 ± 6.8	2.3 ± 2.4	*t*=4.38*
Matrix	10.3 ± 2.7	12.9 ± 2.2	*t*=3.71*
Verbal Reasoning	4.2 ± 1.6	4.7 ± 1.8	*t*=1.10
Cannabis urine status (+ / -)	8 / 19	--	--
Illicit opioid urine status (+ / -)	3 / 24	--	--
Opioid-related medication (buprenorphine, methadone, other[a])[b]	18 / 6 / 3	--	--
Leeds Dependence Scale	20.3 ± 8.7	--	--
Desires for Drug Questionnaire	34.0 ± 19.2	--	--

Notes. Chi-square tests used the likelihood ratio, as a conservative measure; [a]Two participants had been prescribed naltrexone, and another was prescribed gabapentin for managing opioid-related symptoms; [b]OUD participants also took the following prescribed medications at the time of study participation: antidepressants (*N*=12), muscle relaxers (*N*=1), anticonvulsants (*N*=6), stimulants (*N*=4), and benzodiazepines (*N*=5). *$p \leq 0.05$.

Table 2. fMRI metacognition task whole-brain results.

Region	BA	Side	Voxels	Peak T	$P_{cluster}$	x	y	z
Report>Follow: Opioid>Control								
Inf, Sup, Med OFC / Parahippocampus / Ventral and Dorsal Striatum / Gyrus Rectus / sgACC / Olfactory Bulb	38, 28, 25, 11	L	981	5.47	<0.001	-26	16	-24
						-20	2	-26
						4	10	-6
						-10	36	-18
						-14	42	-18
						-10	36	-18
						-8	12	-12
(Negative) Correlation with Report Condition (Parametric Modulation): All Participants								
Mid & Sup Occipital Cortex	18	L	312	5.25	0.003	-24	-92	22
						-12	-92	14
Sup & Inf Parietal Cortex	7	L	460	5.20	<0.001	-16	-70	50
						-32	-56	52
Cerebellum	37	R	220	4.29	0.018	22	-54	-22
						42	-60	-30
Inf Occipital Cortex & Cerebellum	19, 18	L	191	4.20	0.033	-32	-78	-4
						-24	-78	-18

Note. All results were significant at $p<0.05$ cluster-corrected (>180 contiguous voxels), with a search threshold (voxel-wise significance) of $p<0.001$ uncorrected (T>3.25). BA = Brodmann Area, DLPFC = dorsolateral prefrontal cortex, dACC = dorsal anterior cingulate cortex, sgACC = subgenual anterior cingulate cortex, SMA = supplementary motor area, OFC = orbitofrontal cortex, Inf = inferior, Sup = superior, Mid = middle, Med = medial. For results of Report>Follow across all participants, see the Supplement.