Psychological Medicine

Long-term Behavioural Rewriting of Maladaptive Drinking Memories via Reconsolidation-Update Mechanisms.

--Manuscript Draft--

Manuscript Number:	PSM-D-19-01615R2
Full Title:	Long-term Behavioural Rewriting of Maladaptive Drinking Memories via Reconsolidation-Update Mechanisms.
Article Type:	Original Article
Corresponding Author:	Ravi Kumar Das, PhD UCL London, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	UCL
Corresponding Author's Secondary Institution:	
First Author:	Grace Gale, BSc
First Author Secondary Information:	
Order of Authors:	Grace Gale, BSc
	Katie Walsh, PhD
	Vanessa Elizabeth Hennessy, MSc
	Lotte Stemerding, MSc
	Tiffany Koa Sher Ni, MSc
	Emily Thomas, MSc
	Sunjeev K. Kamboj, PhD, DClinPsy
	Ravi Kumar Das, PhD
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Background Alcohol use disorders can be conceptualised as a learned pattern of maladaptive alcohol-consumption behaviours. The memories encoding these behaviours centrally contribute to long-term excessive alcohol consumption and are a key therapeutic target. The transient period of memory instability sparked during memory reconsolidation offers a therapeutic window to directly rewrite these memories using targeted behavioural interventions. However, clinically-relevant demonstrations of the efficacy of this approach are few. We examined key retrieval parameters for destabilising naturalistic drinking memories and the ability of subsequent counterconditioning to effect long-term reductions in drinking. Methods Hazardous/harmful beer-drinking volunteers (N=120) were factorially randomised to retrieve (RET) or not retrieve (No RET) alcohol reward memories with (PE) or without (No PE) alcohol reward prediction error. All participants subsequently underwent disgust-based counterconditioning of drinking cues. Acute responses to alcohol were assessed pre-and post-manipulation and drinking levels assessed up to 9 months. Results

Greater long-term reductions in drinking were found when counterconditioning was conducted following retrieval (with and without PE), despite a lack of short-term group differences in motivational responding to acute alcohol. Large variability in acute levels of learning during counterconditioning were noted. ‘Responsiveness’ to counterconditioning predicted subsequent responses to acute alcohol in RET+PE only, consistent with reconsolidation-update mechanisms.

Conclusions

The longevity of behavioural interventions designed to reduce problematic drinking levels may be enhanced by leveraging reconsolidation-update mechanisms to rewrite maladaptive memory. However, inter-individual variability in levels of corrective learning is likely to determine the efficacy of reconsolidation-updating interventions and should be considered when designing and assessing interventions.

TITLE PAGE:

Long-term Behavioural Rewriting of Maladaptive Drinking Memories via Reconsolidation-Update Mechanisms.

Authors: Gale, Grace., [1] Walsh, Katie., [1] Hennessy, Vanessa E., [1] Stemerding, L.E[1]., Ni, Koa Sher., [1] Thomas, Emily[1] Kamboj, Sunjeev K. [1] & Das, Ravi. K.[1]*

*Corresponding author. All correspondence to Dr. Ravi Das, Clinical, Educational and Health Psychology, UCL, 26 Bedford Way, London WC1H 0AP, United Kingdom, Email: ravi.das@ucl.ac.uk. Telephone 07341311832

Financial Support: This work was funded by a Medical Research Council grant awarded to Sunjeev Kamboj (grant number: MR/M007006/1)

Conflicts of interest: None

Ethical Standards: *The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013.*

WORD COUNT: 4472

ABSTRACT

Background: Alcohol use disorders can be conceptualised as a learned pattern of maladaptive alcohol-consumption behaviours. The memories encoding these behaviours centrally contribute to long-term excessive alcohol consumption and are a key therapeutic target. The transient period of memory instability sparked during memory reconsolidation offers a therapeutic window to directly *rewrite* these memories using targeted behavioural interventions. However, clinically-relevant demonstrations of the efficacy of this approach are few. We examined key retrieval parameters for destabilising naturalistic drinking memories and the ability of subsequent counterconditioning to effect long-term reductions in drinking.

Methods: Hazardous/harmful beer-drinking volunteers (N=120) were factorially randomised to retrieve (RET) or not retrieve (No RET) alcohol reward memories with (PE) or without (No PE) alcohol reward prediction error. All participants subsequently underwent disgust-based *counterconditioning* of drinking cues. Acute responses to alcohol were assessed pre-and post-manipulation and drinking levels assessed up to 9 months.

Results: Greater long-term reductions in drinking were found when counterconditioning was conducted following retrieval (with and without PE), despite a lack of short-term group differences in motivational responding to acute alcohol. Large variability in acute levels of learning during counterconditioning were noted. ‘Responsiveness’ to counterconditioning predicted subsequent responses to acute alcohol in *RET+PE* only, consistent with reconsolidation-update mechanisms.

Conclusions: The longevity of behavioural interventions designed to reduce problematic drinking levels may be enhanced by leveraging reconsolidation-update mechanisms to rewrite

maladaptive memory. However, inter-individual variability in levels of corrective learning is likely to determine the efficacy of reconsolidation-updating interventions and should be considered when designing and assessing interventions.

INTRODUCTION

Harmful drinking and alcohol use disorders (AUDs) represent leading causes of global preventable mortality, contributing to 3 million deaths annually (WHO, 2018) and recent research suggesting an alarming increase in the prevalence of problem drinking in some demographic groups (Grant et al., 2017). Extant treatments for AUD enjoy limited long-term efficacy, with under 20% completing treatment free of dependence and fewer still maintaining abstinence long-term (Public Health England, 2018). Treatment approaches targeting the fundamental processes underlying the development and maintenance of harmful drinking are required to address this global health priority.

AUDs arise via repeated environmental exposure to alcohol amid multivariate risk factors (Sher et al., 2005). Harmful alcohol consumption may therefore be conceptualised partly a *learned* pattern of maladaptive behaviours (Drummond et al., 1990; Hyman, 2005). Alcohol, like other addictive drugs, induces plasticity in mesocorticolimbic motivational circuitry (Pierce & Kumaresan, 2006). This system supports reward learning, adapting behaviour to seek and maximise rewards when environmental cues signal their availability. Alcohol can therefore support behavioural adaptation towards hyper-motivated alcohol seeking and consumption in the presence of environmental 'trigger' cues. Practically, this manifests as arousal, and a strong desire to drink (craving) in response to certain alcohol-predictive contexts and stimuli (e.g. the sight or smell of beer) (Self, 1998; Sinha & Li, 2007).

Memories that support a harmful level of alcohol use, by linking environmental cues to alcohol reward can be considered to be '*maladaptive reward memories*' (MRMs). Once formed through repeated naturalistic exposure to alcohol with accruing drinking episodes (Robbins et al., 2008), these MRMs are highly robust and display remarkable persistence (Hyman & Malenka, 2001) even after extended periods of abstinence. They therefore believed to be a core substrate underlying persistent relapse susceptibility.

Their central pathogenic role suggests MRMs should be a primary target in the treatment of AUDs (Tronson & Taylor, 2013). A novel approach for directly and permanently ameliorating the negative influence of MRMs on behaviour is to leverage the process of memory *reconsolidation* (Milton & Everitt, 2012; Torregrossa & Taylor, 2013). This is a retrieval-dependent memory maintenance process that serves to strengthen and/or update consolidated memory traces when new memory-relevant information is presented at retrieval. Such updating necessitates the temporary *destabilisation* of memory traces, such that new information can be incorporated and the relevant adjustments to the dendritic and synaptic architecture encoding the memory trace made (Clem & Huganir, 2010; Merlo et al., 2015). If adaptive learning (for example, extinction) is timed correctly following retrieval/destabilisation, such that it occurs in the critical (~2 hour) 'reconsolidation window' when memories are active and unstable, it is theoretically possible to *rewrite* maladaptive memory content to a benign form (Germeroth et al., 2017; Monfils & Holmes, 2018). By re-formatting MRMs such that trigger cues do not provoke alcohol seeking, it may be possible to reduce alcohol consumption and prophylactically guard against relapse over the long-term.

Although a nascent field, there are highly promising early demonstrations of the potential of this approach (Walsh et al., 2018). Extinction (i.e. exposure therapy) following retrieval of MRMs has been shown to produce long-lasting reductions in drug-cue-induced craving and physiological arousal (Xue et al., 2012), and reduce smoking in cigarette smokers (Germeroth et al., 2017). However, there have also been notable failures to replicate reconsolidation-interference effects, particularly using the retrieval-extinction paradigm (Baker et al., 2013; Luyten & Beckers, 2017; Soeter & Kindt, 2011). There are several potential reasons for such discrepant results.

Firstly, extinction itself may represent a sub-optimal 'corrective' learning modality, since it is a largely passive procedure, involving no response from participants, unobserved inter-individual variability in engagement and responsiveness to extinction (Shumake et al., 2018) may mask effects. A promising alternative – *counterconditioning*- re-pairs cues reward cues (e.g. pictures of beer) with negatively-valenced outcomes (e.g. disgust-inducing bitter liquids and images). Disgust- counterconditioning may provide a more potent corrective learning experience than extinction (Tunstall et al., 2012) since it 1) leverages a potent food-rejection mechanism (Rozin & Fallon, 1987) 2) the 'disgust' response to certain images and bitter liquids are powerful and virtually universal (Schienle et al., 2015) and 3) it is an 'active' procedure, meaning participants cannot simply disengage from the task, as occurs during extinction. We have shown broad short-term abolition of attentional biases and reactivity to alcohol cues when *counterconditioning* was conducted after MRM retrieval in hazardous drinkers (Das et al., 2015) a finding that has been further demonstrated in experimental animals (Goltseker et al., 2017), however this has never been shown to affect long-term drinking outcomes.

Secondly, memory retrieval and destabilisation are not synonymous. Indeed, memory destabilisation is highly dependent upon various '*boundary condition*s'(Elsey & Kindt, 2017; Walker & Stickgold, 2016). Primary amongst these are the *length* of retrieval (N cues presented), with retrievals that are either too short or too long failing to spark destabilisation (Merlo et al., 2014, 2018; Suzuki et al., 2004) and the presence of an appropriate 'mismatch' learning signal - *prediction error* (PE)(Schultz et al., 1997; Waelti et al., 2001) - at retrieval (Das et al., 2015; Krawczyk et al., 2017; Sevenster et al., 2013). Specifically, some level of mismatch between predicted and actual outcomes is required for destabilisation (Agustina López et al., 2016; Pedreira et al., 2004).

These key parameters have not been systematically manipulated in clinically-focussed reconsolidation interference studies (Walsh et al., 2018). It is unsurprising, then, that findings are inconsistent. In order to properly assess whether rewriting of alcohol MRMs can be reliably achieved through purely behavioural reconsolidation manipulations, systematic investigation of the role of MRM retrieval and prediction error prior to corrective learning is required.

In the current study, we addressed this issue by systematically manipulating MRM retrieval and the presence of prediction error at retrieval prior to a counterconditioning intervention in heavy drinkers. We assess whether the effects of counterconditioning on cue reactivity and drinking levels are potentiated in a retrieval and prediction error-dependent manner, consistent with reconsolidation-based memory rewriting.

METHODS:

Participants & design: 120 hazardous, beer-preferring drinkers were randomised in a 2 (MRM retrieval/ no retrieval) x 2 (prediction error/ no prediction error) factorial design. All participants completed three sessions, corresponding to *baseline* (on Day 1), retrieval/counterconditioning *manipulation* (Day 3-5) and *post-manipulation* (Day 10 – 13). Primary inclusion criteria were : Ages 18-60 , scoring >8 on the Alcohol Use Disorders Identification Test (AUDIT)(Saunders et al., 1993); Consuming > 40 (men) or >30 (women)

UK units/week (1 unit=8g ethanol), drinking ≥4 days each week, primarily drinking beer, and having non-treatment seeking status. Exclusion criteria were: Pregnancy/breastfeeding, diagnosis of AUD/SUDs, current diagnosed psychiatric disorder, AUD as defined by the SCID; use of psychoactive medications, use of illicit drugs > 2x /month.

Measures:

Questionnaire assessments: The comprehensive effects of alcohol questionnaire (CEOA ;Fromme et al., 1993) retrospectively assessed responses to alcohol, the AUDIT, obsessive-compulsive drinking scale (OCDS; Anton et al., 1995) and alcohol craving questionnaire (ACQ-NOW; Singleton et al., 1994) measured maladaptive drinking patterns. Motivation to reduce drinking was measured by the stages of change readiness and treatment eagerness scale (SOCRATES; Miller & Tonigan, 1996). Distress tolerance and sensitivity to disgust were assessed by the Distress Tolerance Scale (DTS; Simons & Gaher, 2005) and Disgust Propensity and Sensitivity Scale (DPSS-R; Olatunji et al., 2007), respectively. Changes in anxiety and affect due to the counterconditioning procedure were assessed using the state version of the Spielberger State-Trait Anxiety Inventory (STAI-S; Spielberger, 2010) and positive and negative affect scale (PANAS; Watson et al., 1988), respectively. Drinking was quantified using the Timeline Follow-Back diary procedure (Sobell & Sobell, 1992). Depressive symptomatology was assessed with the Beck Depression Inventory (BDI)(Beck et al., 1988).

Cue reactivity assessment: As in our previous study (Das et al., 2019), participants were presented with a 150ml glass of beer and told they would consume this after rating a series of images. They then rated their *urge to drink* and *liking of* four ‘orange juice cue’ images and four ‘beer cue’ images. These were subsequently used as retrieval cues in the ‘no retrieval’

(‘No RET’) and retrieval (‘RET’) procedures respectively on the *manipulation* day. Three *wine* and two soft drink (*neutral*) images (not used as retrieval cues) were also rated, followed by *urge to drink* the *in vivo* beer and *predicted enjoyment* of the beer. These were all rated on 11-point (0 to 10) scales. Participants then consumed the beer according to timed on-screen prompts and rated their post-consumption *actual enjoyment* of the beer and *urge to drink more* beer. These scales thus assessed the acute hedonic and motivational properties of alcohol. These *baseline* (*Day 1*) procedures both allowed assessment of changes in cue reactivity and reinforcing properties of alcohol, and set the expectation of beer consumption to maximise PE on the *manipulation* day when the drink was unexpectedly withheld in PE groups during the appropriate retrieval procedure.

MRM retrieval/PE procedure was one we have previously used to reactivate alcohol MRMs and is described fully elsewhere(Das et al., 2015; Das et al., 2019) . Participants’ MRMs were retrieved by viewing/rating beer cues (*RET*). Control memories were retrieved by viewing/rating orange juice cues (*No RET*). This was identical to the cue reactivity task except 1) the *in vivo* beer was replaced with orange juice in the *No RET* groups 2) only four condition-appropriate cue images were rated. To manipulate prediction error (*PE*), the drink given to participants (orange juice or beer) was unexpectedly withheld by an on-screen prompt reading ‘*Stop, do not drink!*’ in *PE* groups: (*RET+PE* and *No RET+PE)* generating negative prediction error. In the ‘*no PE*’ conditions (*RET no PE, No RET no PE*), the drink was consumed as on *Day 1,* as expected.

Counterconditioning: All four groups underwent counterconditioning after the retrieval/PE manipulations as previously described(Das et al., 2018). Briefly, after a 5-minute interval during which participants completed high working memory load distractor tasks (digit span, prose recall), they were shown four beer images and two neutral drink images (coffee and cola) four times each in a pseudo-randomised, fixed order. Two of the beer images (nominated '*Beer-Bit CSs'*) were paired with consumption of 15ml of a highly bitter solution (.067% aqueous Denatonium Benzoate/*Bitrex*). The other two beer images (nominated '*Beer-Pic CSs'*) were followed by one of four images taken from the IAPS database rated highly for induction of disgust. Two coffee and cola images (nominated '*Neut-Neut* CSs') were followed by neutral rated images from the IAPS database. All pairings occurred on a 100% reinforcement ratio. Full information is given in the *supplementary materials*.

Procedure:

Participants responding to study advertisements were screened for eligibility by telephone. On *Day 1*, (*baseline*), participants attended UCL and completed informed consent before being breathalysed (Lion 500 Alcometer) to ensure abstinence from alcohol. They then completed demographic information (gender, age, education and smoking status) and questionnaire measures (AUDIT, Timeline follow-back, OCDS, CEOA, SOCRATES, DTS and BDI). Participants then completed the cue reactivity and acute beer rating, as described above and in the *supplementary materials*.

On *Day 2* (*manipulation: Day 1* + 48-72hrs), breath-alcohol verified abstinence was confirmed prior to completion of the DPSS-R, ACQ-NOW, PANAS and STAI. Participants then underwent group-appropriate retrieval/no-retrieval and PE/no PE manipulation followed by counterconditioning. After completion of counterconditioning participants re-completed the PANAS. On *Day 3* (*post-manipulation:* 7±2 days after *Day 2*) participants attended the test centre for the final time and recompleted all baseline questionnaires and cue reactivity/ acute beer challenge before debriefing.

Remote follow-up assessments of perceived drinking changes, TLFB, ACQ-NOW and SOCRATES measures were completed at 2 weeks, 3, 6 and 9 months following *Day 3*. Participants were reimbursed at the standard university hourly rate (£10) for in-lab testing sessions and incentivised with an extra £5 for each completed remote follow-up.

Sample size was calculated in G*Power 3.1.9.2 for 1-β=.95 to detect a minimum effect size of $n_p{}^2$=.05 at α=.05 for the interaction in a mixed ANOVA, assuming *ρ* of .5. This yielded a total required sample size of N=78 (26 per group). Anticipating minimal attrition, we randomized N=30/group.

Statistical Approach:

See *supplementary materials* for full data-handling. Changes in short-term outcomes (measured in-lab) were assessed with 2 [*Day*: *pre-manipulation* vs. *post-manipulation*) x 2 [*Retrieval: RET* vs *No RET*] x 2 [*PE*: *PE* vs No PE,] mixed ANOVA. For analysis of the cue reactivity, a factor of *Cue Type* (Beer-Bit CS/ Beer-Pic CS/ Neut-Neut CS/Orange Juice/Neutral) was also modelled. For counterconditioning in addition to *RET* and *PE* factors, factors of *Cue Type* (Beer-Bit CS/Beer-Pic CS/Neut-Neut CS) and *Trial* (1st, 2nd, 3rd, final) were included. Where sphericity was violated in repeated measures, the Greenhouse Geisser or multivariate ANOVAs were used, depending on ε values and according to published recommendations(Stevens, 2012). This is reflected in multivariate/non-integer DFs.

Long-term drinking data were analysed using linear mixed models with fixed factors of *Retrieval* and *PE* across *Time* (6:Baseline, Post-manipulation, 2 weeks, 3 months, 6 months, 9 months), modelling per-participant intercepts as baseline values. *Time* slopes were initially modelled as fixed then as random, assessing improvement in model fit according to reduction >2 in Bayesian information criterion (BIC). Due to the presence of highly outlying mean daily unit alcohol consumption values at 2 weeks (~60 units/day, >450/week), an upper-trim on values was performed on means with the trim at 30 units/day. This removed the two outlying data points (males) from the 2-week data, but did not affect other data. Rating data were lost for one participant due to technical error. Alpha for all *a priori* tests was set at .05, with *p*-values Sidak- corrected for post-hoc tests. For tests of baseline trait, drinking and demographics variables, the False Discovery Rate (FDR) correction was applied. Data were analysed blind to condition.

RESULTS:

Participants were largely equivalent at baseline on key variables (see *Table 1*). Due to technical error, post-screening baseline AUDIT data were only available for *No RET no PE* N=22, *No RET+PE* N=20, *RET no PE* N=22, *RET+PE* N=20. There were no differences between groups in number of days between study sessions and this was unrelated to outcomes.

[TABLE 1 HERE]

Counterconditioning: Those in the two retrieval groups were statistically similar in all liking or urge to drink ratings in response to the beer cues and drink used to retrieve MRMs prior to counterconditioning [all Fs (1,58) ≤2.05,*p*s ≥.158]. Inferential statistics for counterconditioning data are given in *Table 2* for clarity. A *Trial*Cue Type* interaction[a] emerged, indicating significant reductions in liking of Bitrex-paired beer CSs[b] and disgust picture-paired beer CSs[c] across trials, with no significant reduction in unreinforced neutral pictures[d]. Counterconditioning thus successfully reduced mean-level *Beer CS* liking. While successful counterconditioning was evident in both *Retrieval* groups, a marginal *Cue Type*Trial*Retrieval* interaction[e] indicated greater liking of *Beer-Bit CSs*[f] and *Neut-Neut CSs*[g] in the *RET* groups vs. *No RET* groups on *Trial 1* of counterconditioning (see *Figure 1*). In the *RET* groups, all *Cue Types* were liked equally on *Trial 1*[h], while in the *No RET* groups liking of *Beer-Pic CSs* was greater than *Neut-Neut CSs*[i] . On *Trial 4* of counterconditioning, *Neut-Neut Css* were liked more than both *Beer* CSs in the *No RET* groups (*p*s≤.014) but not in the *RET* groups (*ps .072 to .956).* Unreinforced pre-exposure to CSs during MRM retrieval may have thus affected the speed and level at which these were differentiated and subsequently counterconditioned as discriminative stimuli. Importantly, however, on *Trial 4*, there were no differences across RET conditions in ratings of cues[j] indicating that absolute responses to counterconditioned cues were similar across groups.

[TABLE 2 HERE]

Counterconditioning response heterogeneity: There was substantial inter-individual variation in ratings of disgust UCSs and CSs across counterconditioning. Descriptive statistics for these ratings are given in *Supplementary Table S2*. Since memory rewriting here is predicated upon level of 'corrective learning' (i.e. effective counterconditioning of beer cues), a measure of '*counterconditioning responsiveness*' was computed as change in liking of CSs across counterconditioning (Trial 4–Trial 1). Greatest variability was seen in ratings of *Beer Pic CSs*. *Responsiveness* was therefore calculated as Trial 4–Trial 1 Δ in *Beer-PIC CS* liking) to be assessed as a predictor in mixed modelling of drinking outcomes and as a covariate where it was correlated with the dependent variable in general linear models (reinforcing effects of beer), including an interaction term with *Group* to assess the difference in the covariate slope across groups. Correlations with key *post-manipulation* outcomes and exploratory analyses of trait predictors of counterconditioning responsiveness are given in *Supplementary Materials* (*Table S3*).

Prediction error generation

Analysis of rated 'surprise' levels following the retrieval and PE/no PE procedures showed a main effect of *PE*, indicating greater surprise in *PE* groups than *no PE* groups [F(1,116) = 309.79, p<.001, η_p^2 = .728]. This did not interact with *Retrieval* group. The PE generation procedure was thus highly successful and equally effective in *RET* and *no RET* groups. Full statistics on manipulation checks for MRM retrieval are given in the *Supplementary materials.*

Primary Outcomes:

Cue reactivity: Reinforcing effects of alcohol

All analyses of reinforcing effects of *in vivo* beer were analysed with *Day* (*baseline* vs. post-manipulation) x *Retrieval* (RET vs. No RET) x *PE* (PE vs. No PE) RMANCOVAs, including counterconditioning *Responsiveness* as a covariate that could interact with *RET*PE*. Four-way interactions were found for pre-consumption *anticipated enjoyment* and *urge to drink* beer and post-consumption (primed) *urge to drink more* beer. Commensurate with the bivariate correlations, the 4-way interactions were driven *Day*Responsiveness* interactions in *RET+PE* only, indicating that degree of achieved counterconditioning predicted *post-manipulation* reactivity to *in-vivo* beer only in the 'active' *RET+PE* group. For *actual enjoyment* of beer (post consumption), counterconditioning responsiveness again predicted *post-manipulation* enjoyment only in *RET+* However, the 4-way interaction did not reach significance. These interaction terms and simple slopes are given in *Table 3*. Scatterplots of bivariate associations are given in *Figure 2*. Analysis of ratings of pictorial cues used in the cue reactivity task are given in the *supplementary materials*.

[TABLE 3 HERE]

[FIGURE 2 HERE]

Drinking levels:

Beer

The random intercepts-only effects mixed model revealed a significant main effect of *Time* [F(1,522.74)=39.027, *p*<.001] and a marginally significant *RET*PE*Time* interaction [F(1,522.74)=3.965, *p*=.047]. The *Time* effect represented a reduction in beer consumption across the follow-up period, with a mean reduction of .23 UK pints/day at each time point [*b*=-.232, *t*(521.5)=2.04, *p*<.0005]. The 3-way interaction represented a greater reduction in drinking across *Time* in *RET+PE* than *No RET+PE* [*b*=.146, *t*=2.06, *p*=.0397], with no differences between the other groups. Model-predicted and true values for this effect are shown below in *Figure 3* panels A and B. Modelling random slopes for *Time* did not improve model fit (BIC 2128.485→2128.919) and yielded non-significant variance in slopes (Z=1.138, *p*=.255). *Responsiveness* to counterconditioning was not a significant predictor [F(1,119.495)=.72, *p*=.679] and was detrimental to parsimonious model fit (BIC 2128.485→2134.752).

Total Units

The random intercepts-only model for total unit consumption data (BIC=3748.009) yielded a significant effect of *Time* [F(1,533.775)=25.487, *p*<.001] and *RET*Time* interaction [F(1, 533.775)=4.937, *p*=.027]. Simple contrasts on the *Time* main effect against baseline drinking levels showed no overall change in drinking from baseline to post-manipulation [*b*=-.69, *t*(511.97)=.706, *p*=.48] or 2 weeks [*b*=-1.196, *t*(516.53)=.1.194, *p*=.233], with a marginal reduction by 3 months [*b*=-1.97, *t*(519.482)=1.925, *p*=.055] and significant reductions by 6 months [*b*=-4.66, *t*(519.48)=4.549, *p*<.001] and 9 months [*b*=-3.65, *t*(521.05)=3.431, *p*=.001]. Parameter estimates for the *RET*Time* interaction showed a greater reduction in drinking across *Time* in *RET* than *No RET* groups [*b*=.575, *t*(531.58)=2.192, *p*=.029]. Within-groups, the slope for the reduction in drinking across time was highly significant in the *RET* groups [*b*=-.923, *t*(51.26)=-5.008, *p*<.0005] but non-significant in the *No RET* groups [*b*=-.3, *t*(53.958)=-1.177, *p*=.245].

Significant variance in slopes [Z=2.781, *p*=.005] and improved model fit [Δ-2LL $\chi^2(2)$=-18.004, *p* <.001, BIC 3748.09→3743.262] when allowing slopes for *Time* to vary indicated that a random slopes effect model was appropriate. This reduced the *RET*Time* effect to only a marginally significant level [*b*=.623, *t*(107.023)=1.999, *p*=.049]. Including counterconditioning *Responsiveness* as a covariate yielded a borderline-significant predictive impact in drinking [F(1,119.518)=3.916, *p*=.05], but was detrimental to parsimonious model fit [3743.262 →3749.194], so was not included in the final model. Actual and mean model-predicted values for the *RET*Time* effect in the final model are shown in *Figure 3* panels C&D.

[FIGURE 3 HERE]

DISCUSSION

We examined the potential for putative memory reconsolidation mechanisms to catalyse the efficacy and longevity of an experimental learning-based intervention in ameliorating maladaptive drinking patterns. We found mixed evidence that supported the long-term utility of a reconsolidation-focussed approach, while highlighting large response variability and potential limitations of a homogenous learning manipulation.

We observed a greater reduction in over the 9 months follow-up period when counterconditioning followed the -putatively 'active' *retrieval* (*RET*) *with prediction error* (PE) manipulation. Greater reductions in non-specific, *total* alcohol consumption were seen in both MRM retrieval groups, although this was not PE-dependent. These results are broadly consistent with counterconditioning updating MRMs via reconsolidation mechanisms, producing lasting beneficial changes in drinking behaviour. That lasting effects on drinking levels are observed after a one-off, purely behavioural manipulation is encouraging and extends our previous work on ketamine, suggesting reconsolidation-focussed therapies may have a bright future in the treatment of SUDs.

The current results extend our previous findings with counterconditioning during the reconsolidation window (Das et al., 2015) and pharmacological blockade of alcohol MRM reconsolidation by ketamine (Das et al., 2019) . While we previously demonstrated *RET* and *PE* –dependent beneficial effects of counterconditioning on computerised in-lab markers of MRMs, changes in responses to actual alcohol and long-term reductions in drinking following have not, until now, been shown using a purely behavioural reconsolidation-update manipulation.

Unexpectedly, the beneficial effects observed here were primarily evident only in the longer-term drinking outcomes but not acute in-lab measures of cue reactivity. The reason for this discrepancy is uncertain. One possibility is lack of sensitivity or limited ecological validity of an in-lab acute assessment of the reinforcing effects of alcohol, since anticipated enjoyment and urge to drink have no impact on whether beer is consumed or not during this test. An emergent and more compelling interpretation is that memory rewriting manipulations display their true utility when participants are exposed to naturalistic ‘high-risk’ relapse scenarios following manipulation. Indeed, previous research has also observed lagged improvements in phobic symptomatology (Soeter & Kindt, 2015) and craving reductions and CO levels in smokers (Germeroth et al., 2017) following a reconsolidation intervention. This is in line with protection against renewal, reinstatement and spontaneous recovery conferred by reconsolidation interference in the experimental literature. The follow-up period used here is the longest of which we are aware in the reconsolidation literature and the potential for these lagged effects highlights the importance of assessing the longevity of effects over extended follow-up.

Short-term improvements are typically seen following learning-based interventions such as cue-exposure therapy, but these are not maintained across time and contexts. Indeed, in the current study, all groups largely displayed improvements in maladaptive drinking behaviours from pre–to-post-manipulation. Incorporating prior retrieval/destabilisation of MRMs offers a potential means to make these interventions '*stick*', vastly enhancing their long-term efficacy and protecting against relapse. The 'single-shot' nature of reconsolidation-interference means it could readily be included as part of a comprehensive psychological treatment program with minimal addition to therapist/patient burden. It may potentially act synergistically with other treatment components that target the biological, cognitive and social causes of AUD by addressing a core, low-level relapsogenic mechanism.

The discrepancy between retrieval and prediction-error-dependent effects on beer vs. all alcohol consumption was unexpected. We and others (Agustina López et al., 2016; Das et al., 2015; Exton-McGuinness et al., 2015; Krawczyk et al., 2017; Sevenster et al., 2014) have previously forwarded PE or 'surprise' at retrieval as a necessary condition for destabilisation of consolidated memories. Hypothetically, PE signals insufficient or inaccurate prediction of outcomes currently stored by the memory trace and necessitates memory destabilisation to allow the memory to update and stay 'relevant'. These findings may seem to suggest that PE is of secondary importance in sparking memory destabilisation and reconsolidation. Indeed, most previous experimental (Milton et al., 2008; Monfils & Holmes, 2018; Saitoh et al., 2017) and clinically applied (Germeroth et al., 2017; Xue et al., 2012, 2017) reconsolidation studies reporting positive findings have not explicitly manipulated PE. There are several key points that should be borne in mind which caution against such an interpretation, however.

It is typical in reconsolidation studies to omit the primary reinforcer during cue-driven retrieval. This will generate a variable level of PE to the extent that reinforcement is expected, despite not explicitly aiming to manipulate PE. In clinical populations, where craving/desire to use is likely to be high to response to drug cues, we may reasonably expect greater PE when drug is not consumed. This is supported by the association between anticipated liking and urge to drink observed and subsequent PE seen in the current study (see Supplementary Materials). This may well account for variability in previous findings. In the current study, although not statistically significant, the *RET+PE* group also showed the steepest overall absolute decrease in overall drinking, meaning unintended PE generation in the *RET no PE* group may have limited power to observe PE-dependent effects. Indeed, peri-retrieval '*surprise*' ratings demonstrated some variability in surprise in the *RET no PE* and *RET+PE* groups, indicating that some level of unintended PE was occurring in the former group and some expectancy of deception in the latter. For clinical translation, there is minimal extra burden involved in explicitly generating and assessing PE during MRM retrieval. Indeed, in treatment scenarios (e.g. in detoxified drug-abusing patients) it would be ethically unacceptable to reinforce patients with abused drugs. Moreover, there are no demonstrations of *inferiority* of PE vs. no PE at retrieval in memory destabilisation, thus the most prudent course of action would be to include PE-generation procedures in experimental and translational retrieval protocols going forward and at the very least assess these explicitly. As a minimum criterion, 'reactivation' cues should evoke an urge/desire to consume and anticipatory enjoyment of drug reward. These measures may be predictive of outcome variability where PE is not assessed.

Limitations:

We have previously assumed a relatively homogenous response to the counterconditioning intervention, given that is leverages very basic learning and aversion mechanisms. The large observed variability in level of achieved counterconditioning or 'responsiveness' demonstrate that this assumption is not tenable. Some participants displayed reductions of in liking of negatively reinforced beer stimuli over half the scale range while others showed little or no change and some even displayed *increased* liking over the course of the task. Equally, some participants did not rate the UCSs as particularly aversive, with some even rating them as mildly pleasant. Having extensively piloted the doses of Bitrex used here ourselves, this is puzzling to us, although genetic polymorphisms moderating bitterness perception may play a key role (Duffy & Bartoshuk, 2000). We further found that disgust propensity, sensitivity and distress tolerance predicted counterconditioning responsiveness, yielding potentially useful trait markers of likely treatment response. However, such individual variability to counterconditioning likely obscured potential group-level differences in responses to the acute alcohol challenge. Interestingly, the 'degree' of counterconditioning was predictive of proximal markers of responding to alcohol, but not long-term drinking outcomes. We believe this is a largely statistical phenomenon, due to greater variance in drinking levels vs. in-lab measures of cue reactivity. However, it is possible that with passing time since reconsolidation-intervention and possible 'schematisation' of updated associations, the degree of acute 'responsiveness' to counterconditioning becomes less critical to outcomes. This would needto validated empirically, but further highlights a potential disparity between proximal and enduring measures of intervention response and underscores the importance of long-term follow-up.

One could reasonably anticipate equal (or greater) response variability when using retrieval-extinction (Shumake et al., 2018); a paradigm that has dominated behavioural memory rewriting research. This may partially explain the inconsistencies and difficulties replicating findings with retrieval-extinction interventions (Baker et al., 2013; Chen et al., 2014; Luyten & Beckers, 2017; Soeter & Kindt, 2011), since a failure to extinguish would preclude any potentiating effect of prior memory retrieval. These observations highlight the importance assessing level of corrective learning, conducting learning to a criterion level or identifying potential low-responders within reconsolidation-updating paradigms.

Variability in learning is perhaps a reason to recommend pharmacological memory-weakening over purely behavioural memory updating approaches in certain populations. Drugs' pharmacodynamic profiles are generally not subject to influence by individual cognitive variables like learning rates, boredom and punishment insensitivity and may be a key option where behavioural approaches fail.

There is no way of assessing whether the *RET+PE* truly destabilised alcohol MRMs and engaged reconsolidation mechanisms (or did so to an equal degree) in all individuals in the current study, since memory destabilisation is a behaviourally silent process. This remains the primary impediment to translational/clinical developments within the reconsolidation field, which is in desperate need of validated biomarkers of memory destabilisation. The lack of triangulation between short-term lab measures and longer-term drinking outcomes compounds this issue in the current study. We have, however, now demonstrated group-level sufficiency of the *RET+PE* procedure used improving clinically-relevant outcomes in five studies (Das et al., 2018; Das et al., 2015; Das et al., 2018, 2019; Hon et al., 2016). Along with the apparently durable effects on drinking observed here, this lends support to the notion that reconsolidation mechanisms were engaged in the current study. While non reconsolidation mechanisms may explain shorter-term effects on outcome, the emergence of divergent effects longer-term observed here are in line with reconsolidation-update.

The current study highlights fundamental questions regarding the parameters that conspire retrieval conspire to determine the fate of memories at retrieval. The future of memory-rewriting interventions will rely upon better understanding of these parameters and individual optimisation of memory destabilisation procedures based therein. Nevertheless, the results obtained here are should energise future research in the field, particularly to assess whether similar effects can be replicated in clinically diagnosed samples where comorbidities and cognitive impairment from chronic alcohol abuse may further complicate implementation.

TABLE AND FIGURE LEGENDS:

Table 1: Baseline demographic drinking and questionnaire measures. Groups did not differ at false-discovery rate (FDR)-corrected alpha for any variables at baseline. Degrees of freedom for one-way ANOVA are all 3, 116, with the exception of AUDIT data where DFs were 1,83 due to data loss.

Table 2: Key inferential statistics for cue liking data during the counterconditioning task. Higher-order effects are given in bold, with the simple-effects analyses used to unpick interactions beneath. Beer-Bit CSs= beer cues paired with Bitrex. *Beer-Pic CSs* = Beer cues paired with disgust images, *Neut-Neut CSs* = Neutral images paired with neutral images (control). Superscript letters refer to the terms discussed in the text.

Table 3: Reactivity to in-vivo beer: Highest-order (four-way) interaction terms in Day*Retrieval*Responsiveness*PE mixed ANOVAs on anticipated and actual enjoyment of sampled beer and pre and post-drink urge to drink beer. Significant effects are highlighted in bold. Degrees of freedom (DFs)=29 for all t-tests.

Figure 1: Liking ratings for the conditioned stimuli across the counterconditioning task. Significant reductions in liking of the Bitrex-paired beer CS (*Beer-Bit CS*) and disgusting image-paired beer CS (*Beer-Pic CS*) were seen in reactivated and non-reactivated groups. However only in *No RET* did the liking of CSs differ on Trial 1. *****=*Beer-Pic>Neut-Neut,* **¥**=*Beer-Pic>Beer-Bit,* **†**=*Neut-Neut>Beer-Bit,* **#**=*Neut-Neut>Beer-Pic.*

Figure 2: Associations between 'strength' of counterconditioning (change in liking of counterconditioned beer cues) anticipated enjoyment, urge to drink, actual enjoyment and urge to drink more beer on the Day 3 beer reactivity test. The correlations were significant only in RET+PE (rightmost column). Dashed lines are ordinary least-squares linear best fit lines.

Figure 3: **Panel A** (top left) changes in mean daily beer consumption (in UK pints) across the study time points in each group. **Panel B (top right)** Mixed model fit values for beer consumption data. A marginally significant *Time*RET*PE* interaction indicated a steeper reduction across Time in *RET+PE* than *No RET+PE* (p=.037). **Panel C***:* Changes in mean daily unit alcohol consumption across the study time points in each group. **Panel D***:* Model fit values for overall alcohol consumption (total UK unit) data. A significant *RET*Time* interaction indicated significant reductions across time in RET groups but not No RET groups. Panels A&C, error bars represent SD. Panels B and D, error bars represent model SEMs.

REFERENCES

Agustina López, M., Jimena Santos, M., Cortasa, S., Fernández, R. S., Carbó Tano, M., & Pedreira, M. E. (2016). Different dimensions of the prediction error as a decisive factor for the triggering of the reconsolidation process. *Neurobiology of Learning and Memory, 136*, 210–219. https://doi.org/10.1016/j.nlm.2016.10.016

Anton, R. F., Moak, D. H., & Latham, P. (1995). The Obsessive Compulsive Drinking Scale: A Self-Rated Instrument for the Quantification of Thoughts about Alcohol and Drinking Behavior. *Alcoholism: Clinical and Experimental Research*, *19*(1), 92–99. https://doi.org/10.1111/j.1530-0277.1995.tb01475.x

Baker, K. D., McNally, G. P., & Richardson, R. (2013). Memory retrieval before or after extinction reduces recovery of fear in adolescent rats. *Learning and Memory*. https://doi.org/10.1101/lm.031989.113

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*(1), 77–100.

Chen, S., Cai, D., Pearce, K., Sun, P. Y. W., Roberts, A. C., & Glanzman, D. L. (2014). Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in Aplysia. *ELife*, *3*, e03896. https://doi.org/10.7554/eLife.03896

Clem, R. L., & Huganir, R. L. (2010). Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. *Science, 330*(6007), 1108–1112.

Das, R. K., Gale, G., Walsh, K., Hennessy, V. E., Iskandar, G., Mordecai, L. A., Brandner, B., Kindt, M., Curran, H. V., & Kamboj, S. K. (2019). Ketamine can reduce harmful drinking by pharmacologically rewriting drinking memories. *Nature Communications*, *10*(1), 5187. https://doi.org/10.1038/s41467-019-13162-w

Das, R.K., Lawn, W., & Kamboj, S. K. (2015). Rewriting the valuation and salience of alcohol-related stimuli via memory reconsolidation. *Translational Psychiatry*, *5*(9), e645–e645. https://doi.org/10.1038/tp.2015.132

Das, R.K., Walsh, K., Hannaford, J., Lazzarino, A. I., & Kamboj, S. K. (2018). Nitrous oxide may interfere with the reconsolidation of drinking memories in hazardous drinkers in a prediction-error-dependent manner. *European Neuropsychopharmacology*, *28*(7), 828–840. https://doi.org/10.1016/j.euroneuro.2018.05.001

Das, R.K., Gale, G., Hennessy, V., & Kamboj, S. K. (2018). A Prediction Error-driven Retrieval Procedure for Destabilizing and Rewriting Maladaptive Reward Memories in Hazardous Drinkers. *Journal of Visualized Experiments*, *56097*(131), e56097–e56097. https://doi.org/10.3791/56097

Drummond, D. C., Cooper, T., & Glautier, S. P. (1990). Conditioned learning in alcohol dependence: implications for cue exposure treatment. *British Journal of Addiction*, *85*(6), 725–743.

Duffy, V. B., & Bartoshuk, L. M. (2000). Food acceptance and genetic variation in taste. *Journal of the American Dietetic Association*, *100*(6), 647–655. https://doi.org/10.1016/S0002-8223(00)00191-7

Elsey, J. W. B., & Kindt, M. (2017). Breaking boundaries: optimizing reconsolidation-based interventions for strong and old memories. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *24*(9), 472–479. https://doi.org/10.1101/lm.044156.116

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories-The role of prediction errors in memory reconsolidation. In *Behavioural Brain Research* (Vol. 278). https://doi.org/10.1016/j.bbr.2014.10.011

Fromme, K., Stroot, E. A., & Kaplan, D. (1993). Comprehensive effects of alcohol: Development and psychometric assessment of a new expectancy questionnaire. *Psychological Assessment*, *5*(1), 19–26. https://doi.org/10.1037/1040-3590.5.1.19

Germeroth, L. J., Carpenter, M. J., Baker, N. L., Froeliger, B., LaRowe, S. D., & Saladin, M. E. (2017). Effect of a Brief Memory Updating Intervention on Smoking Behavior. *JAMA Psychiatry*, *74*(3), 214. https://doi.org/10.1001/jamapsychiatry.2016.3148

Goltseker, K., Bolotin, L., & Barak, S. (2017). Counterconditioning During Reconsolidation Prevents Relapse of Cocaine Memories. *Neuropsychopharmacology*, *42*(3), 716–726. https://doi.org/10.1038/npp.2016.140

Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., Huang, B., Jung, J., Zhang, H., Fan, A., & Hasin, D. S. (2017). Prevalence of 12-Month Alcohol Use, High-Risk Drinking, and DSM-IV Alcohol Use Disorder in the United States, 2001-2002 to 2012-2013. *JAMA Psychiatry*, *74*(9), 911. https://doi.org/10.1001/jamapsychiatry.2017.2161

Hon, T., Das, R. K. R. K., & Kamboj, S. K. S. K. S. K. (2016). The effects of cognitive reappraisal following retrieval-procedures designed to destabilize alcohol memories in high-risk drinkers. *Psychopharmacology*, *233*(5), 851–861. https://doi.org/10.1007/s00213-015-4164-y

Hyman, S. E. (2005). Addiction: a disease of learning and memory. *American Journal of Psychiatry*, *162*(8), 1414–1422.

Hyman, S. E., & Malenka, R. C. (2001). Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature Reviews Neuroscience*, *2*(10), 695–703.

Krawczyk, M. C., Fernández, R. S., Pedreira, M. E., & Boccia, M. M. (2017). Toward a better understanding on the role of prediction error on memory processes: From bench to clinic. *Neurobiology of Learning and Memory*, *142*(Part A), 13–20. https://doi.org/10.1016/j.nlm.2016.12.011

Luyten, L., & Beckers, T. (2017). A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. *Neurobiology of Learning and Memory*. https://doi.org/10.1016/j.nlm.2017.07.014

Merlo, E., Bekinschtein, P., Jonkman, S., & Medina, J. H. (2015). Molecular Mechanisms of Memory Consolidation, Reconsolidation, and Persistence. *Neural Plasticity*, *2015*, 1–2. https://doi.org/10.1155/2015/687175

Merlo, E., Milton, A. L., & Everitt, B. J. (2018). A Novel Retrieval-Dependent Memory Process Revealed by the Arrest of ERK1/2 Activation in the Basolateral Amygdala. *The Journal of Neuroscience*, *38*(13), 3199–3207. https://doi.org/10.1523/JNEUROSCI.3273-17.2018

Merlo, E., Milton, A. L., Goozee, Z. Y., Theobald, D. E., & Everitt, B. J. (2014). Reconsolidation and Extinction Are Dissociable and Mutually Exclusive Processes: Behavioral and Molecular Evidence. *Journal of Neuroscience*, *34*(7), 2422–2431. https://doi.org/10.1523/JNEUROSCI.4001-13.2014

Miller, W. R., & Tonigan, J. S. (1996). Assessing drinkers’ motivation for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*, *10*(2), 81–89. https://doi.org/10.1037/0893-164X.10.2.81

Milton, A. L., & Everitt, B. J. (2012). The persistence of maladaptive memory: addiction, drug memories and anti-relapse treatments. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1119–1139. https://doi.org/10.1016/j.neubiorev.2012.01.002

Milton, A. L., Lee, J. L. C., Butler, V. J., Gardner, R., & Everitt, B. J. (2008). Intra-amygdala and systemic antagonism of NMDA receptors prevents the reconsolidation of drug-associated memory and impairs subsequently both novel and previously acquired drug-seeking behaviors. *The Journal of Neuroscience*, *28*(33), 8230–8237.

Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*, *5*(12), 1032–1042. https://doi.org/http://dx.doi.org/10.1016/S2215-

0366%2818%2930270-0

Olatunji, B. O., Cisler, J. M., Deacon, B. J., Connolly, K., & Lohr, J. M. (2007). The Disgust Propensity and Sensitivity Scale-Revised: Psychometric properties and specificity in relation to anxiety disorder symptoms. *Journal of Anxiety Disorders*, *21*(7), 918–930. https://doi.org/10.1016/j.janxdis.2006.12.005

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory*, *11*(5), 579–585.

Pierce, R. C., & Kumaresan, V. (2006). The mesolimbic dopamine system: The final common pathway for the reinforcing effect of drugs of abuse? *Neuroscience & Biobehavioral Reviews*, *30*(2), 215–238. https://doi.org/http://dx.doi.org/10.1016/j.neubiorev.2005.04.016

Public HealthEngland, Department of Health, & National Drug Evidence Centre. (2018). *Adult Drug Statistics from the National Drug Treatment Monitoring System (NDTMS). April 2017*, 38. www.facebook.com/PublicHealthEngland

Robbins, T. W., Ersche, K. D., & Everitt, B. J. (2008). Drug addiction and the memory systems of the brain. *Annals of the New York Academy of Sciences*, *1141*(1), 1–21.

Rozin, P., & Fallon, A. E. (1987). A Perspective on Disgust. *Psychological Review*. https://doi.org/10.1037/0033-295X.94.1.23

Saitoh, A., Akagi, K., Oka, J.-I., & Yamada, M. (2017). Post-reexposure administration of d-cycloserine facilitates reconsolidation of contextual conditioned fear memory in rats. *Journal of Neural Transmission*, *124*(5), 583–587. https://doi.org/10.1007/s00702-017-1704-0

Saunders, J. B., Aasland, O. G., Babor, T. F., Delafuente, J. R., Grant, M., De La Fuente, J. R., Grant, M., Delafuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schienle, A., Arendasy, M., & Schwab, D. (2015). Disgust Responses to Bitter Compounds: the Role of Disgust Sensitivity. *Chemosensory Perception*, *8*(4), 167–173. https://doi.org/10.1007/s12078-015-9186-7

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Self, D. W. (1998). Neural substrates of drug craving and relapse in drug addiction. *Annals of Medicine*, *30*(4), 379–389. https://doi.org/10.3109/07853899809029938

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science*, *339*(6121), 830–833. https://doi.org/10.1126/science.1231357

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory*, *21*(11), 580–584. https://doi.org/10.1101/lm.035493.114

Sher, K. J., Grekin, E. R., & Williams, N. A. (2005). The Development of Alcohol Use Disorders. *Annual Review of Clinical Psychology*, *1*(1), 493–523. https://doi.org/10.1146/annurev.clinpsy.1.102803.144107

Shumake, J., Jones, C., Auchter, A., & Monfils, M.-H. (2018). Data-driven criteria to assess fear remission and phenotypic variability of extinction in rats. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1742), 20170035. https://doi.org/10.1098/rstb.2017.0035

Simons, J. S., & Gaher, R. M. (2005). The Distress Tolerance Scale: Development and Validation of a Self-Report Measure. *Motivation and Emotion*, *29*(2), 83–102.

https://doi.org/10.1007/s11031-005-7955-3

Singleton, E. G., Henningfield, J. E., & Tiffany, S. T. (1994). Alcohol craving questionnaire: ACQ-Now: background and administration manual. *Baltimore: NIDA Addiction Research Centre*.

Sinha, R., & Li, C. S. R. (2007). Imaging stress- and cue-induced drug and alcohol craving: association with relapse and clinical implications. *Drug and Alcohol Review*, *26*(1), 25–31. https://doi.org/10.1080/09595230601036960

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back. In *Measuring alcohol consumption* (pp. 41–72). Springer.

Soeter, M., & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. *Learning & Memory*, *18*(6), 357–366. https://doi.org/10.1101/lm.2148511

Soeter, M., & Kindt, M. (2015). An abrupt transformation of phobic behavior after a post-retrieval amnesic agent. *Biological Psychiatry*, *78*(12), 880–886. https://doi.org/10.1016/j.biopsych.2015.04.006

Spielberger, C. D. (2010). *State‐ Trait Anxiety Inventory*. Wiley Online Library.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *The Journal of Neuroscience*, *24*(20), 4787–4795.

Torregrossa, M. M., & Taylor, J. R. (2013). Learning to forget: manipulating extinction and reconsolidation processes to treat addiction. *Psychopharmacology*, *226*(4), 659–672.

Tronson, N. C., & Taylor, J. R. (2013). Addiction: A drug-induced disorder of memory reconsolidation. In *Current Opinion in Neurobiology* (Vol. 23, Issue 4, pp. 573–580). Elsevier Current Trends. https://doi.org/10.1016/j.conb.2013.01.022

Tunstall, B. J., Verendeev, A., & Kearns, D. N. (2012). A comparison of therapies for the treatment of drug cues: Counterconditioning vs. extinction in male rats. *Experimental and Clinical Psychopharmacology*, *20*(6), 447–453. https://doi.org/10.1037/a0030593

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48.

Walker, M. P., & Stickgold, R. (2016). Understanding the boundary conditions of memory reconsolidation. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 113, Issue 28, pp. E3991–E3992). https://doi.org/10.1073/pnas.1607964113

Walsh, K. H., Das, R. K., Saladin, M. E., & Kamboj, S. K. (2018). Modulation of naturalistic maladaptive memories using behavioural and pharmacological reconsolidation-interfering strategies: a systematic review and meta-analysis of clinical and 'sub-clinical' studies. *Psychopharmacology*, *235*(9), 2507–2527. https://doi.org/10.1007/s00213-018-4983-8

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

WHO | Global status report on alcohol and health. (2018). *WHO*. https://www.who.int/substance_abuse/publications/global_alcohol_report/en/

Xue, Y.-X., Deng, J.-H., Chen, Y.-Y., Zhang, L.-B., Wu, P., Huang, G.-D., Luo, Y.-X., Bao, Y.-P., Wang, Y.-M., Shaham, Y., Shi, J., & Lu, L. (2017). Effect of selective inhibition of reactivated nicotine-associated memories with propranolol on nicotine craving. *JAMA Psychiatry*, *74*(3), 224–232. https://doi.org/10.1001/jamapsychiatry.2016.3907

Xue, Y.-X., Luo, Y.-X., Wu, P., Shi, H.-S. H.-S., Xue, L.-F., Chen, C., Zhu, W.-L., Ding, Z.-B., Bao, Y., Shi, J., Epstein, D. H., Shaham, Y., & Lu, L. (2012). A Memory Retrieval-

Extinction Procedure to Prevent Drug Craving and Relapse. *Science*, *336*(6078), 241–245. https://doi.org/10.1126/science.1215070

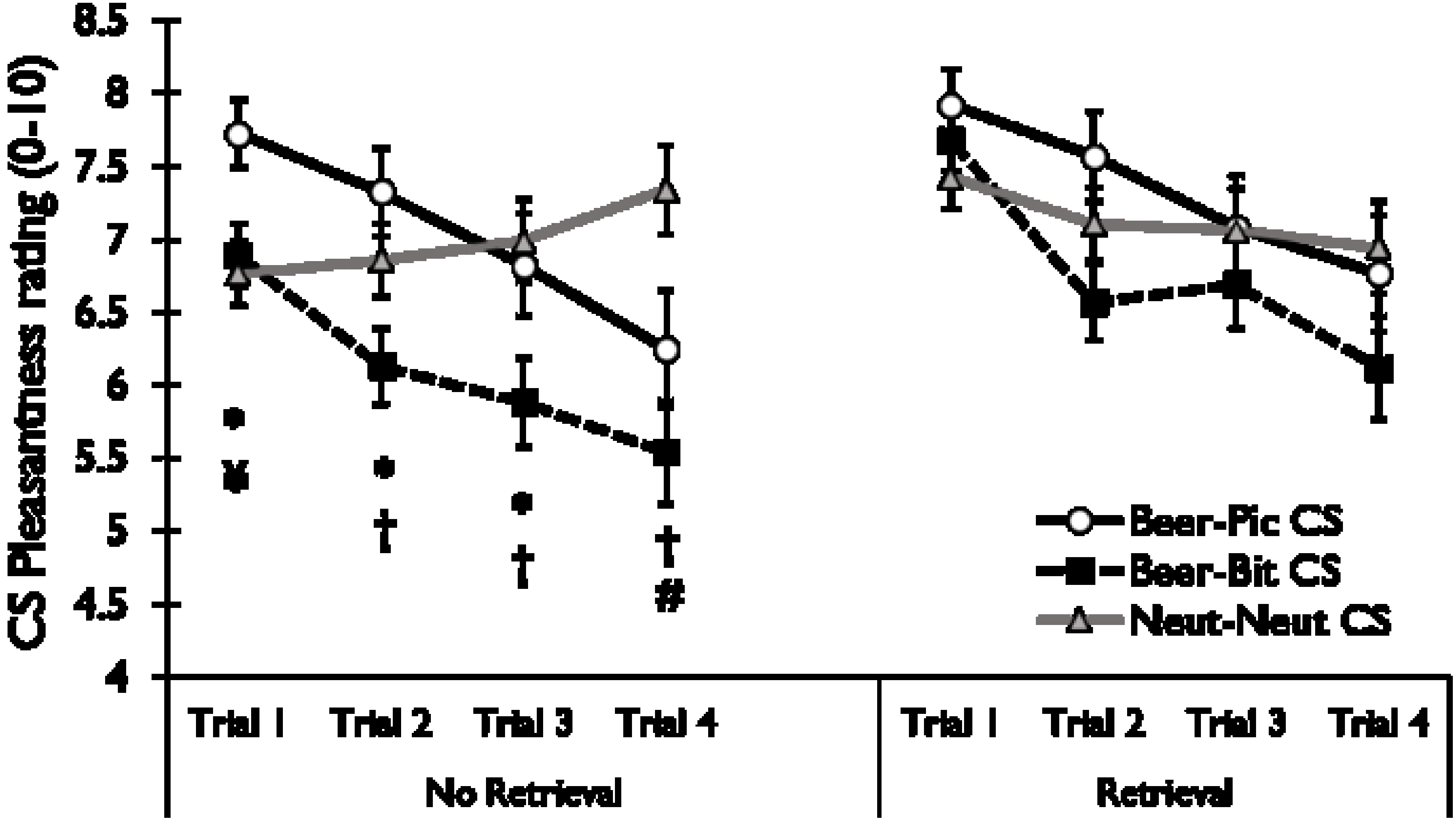

CS Pleasantness rating (0-10)
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
*
¥
*
†
*
†
†
#
Beer-Pic CS
Beer-Bit CS
Neut-Neut CS
Trial 1
Trial 2
Trial 3
Trial 4
No Retrieval
Trial 1
Trial 2
Trial 3
Trial 4
Retrieval

Figure 2

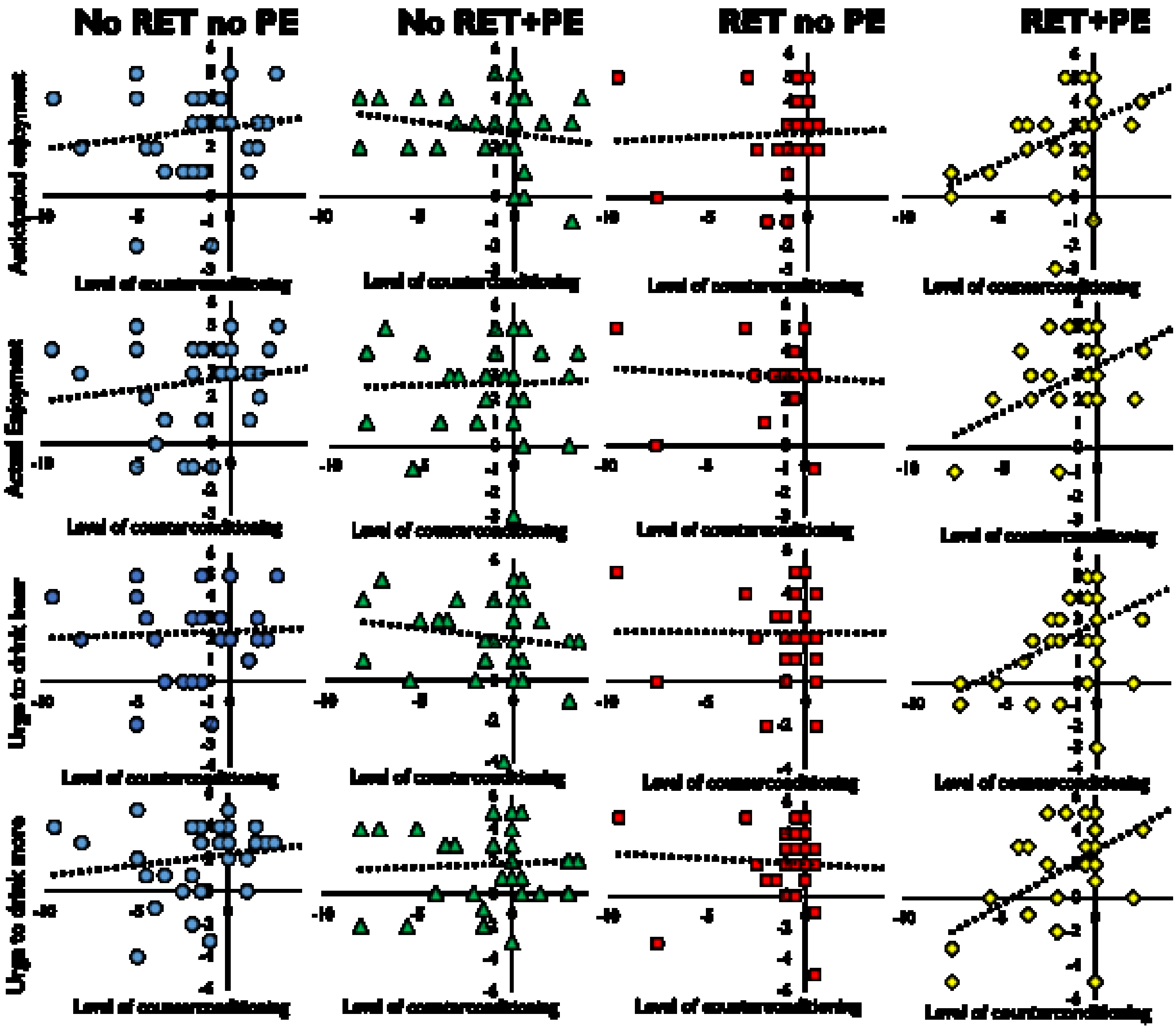
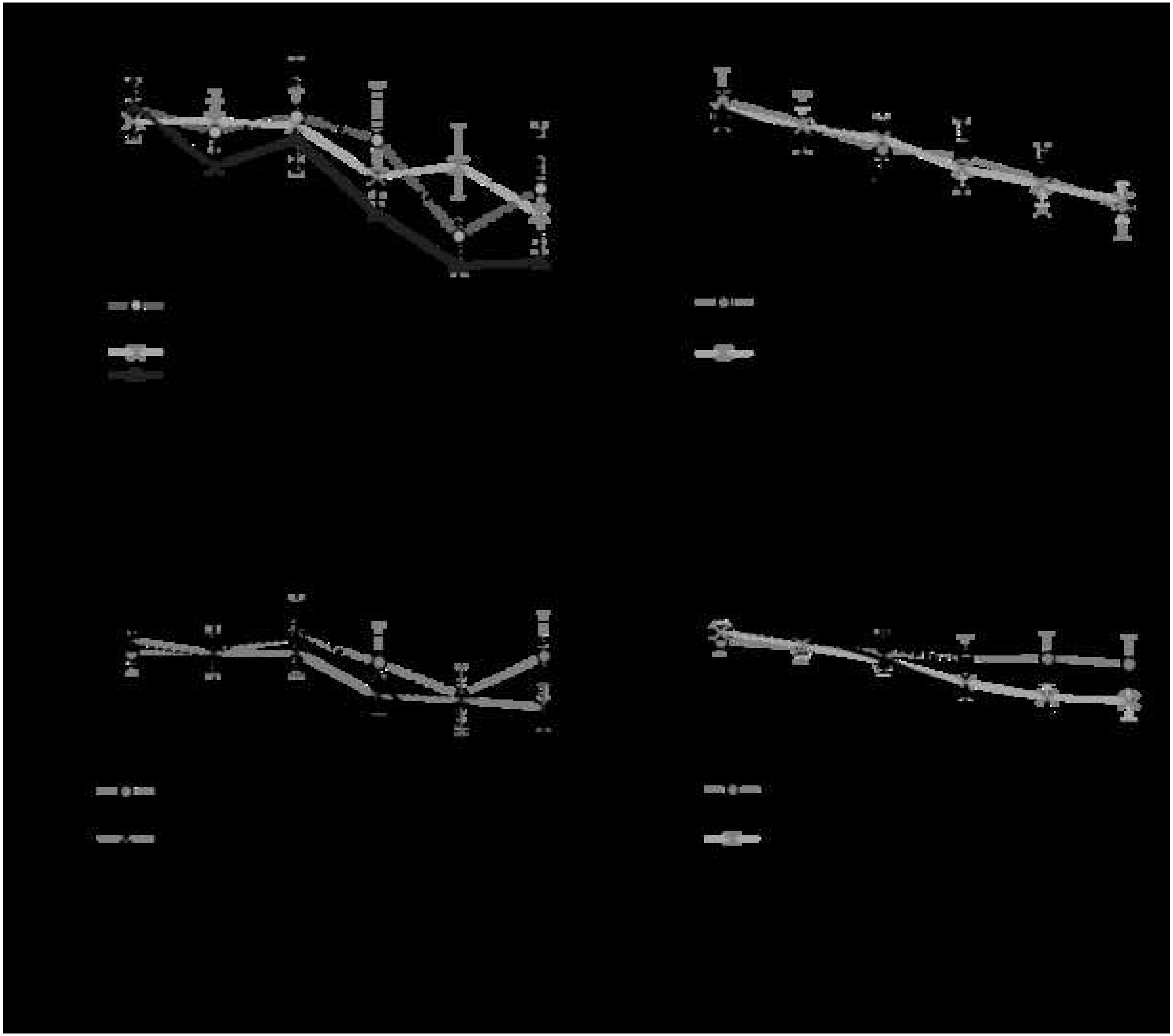

No RET no PE
No RET+PE
RET no PE
RET+PE
Anticipated enjoyment
Actual Enjoyment
Urge to drink beer
Urge to drink more
Level of counterconditioning
-10
-5
6
5
4
3
2
1
0
-1
-2
-3
-4

Figure 3



Supplementary Figure S1

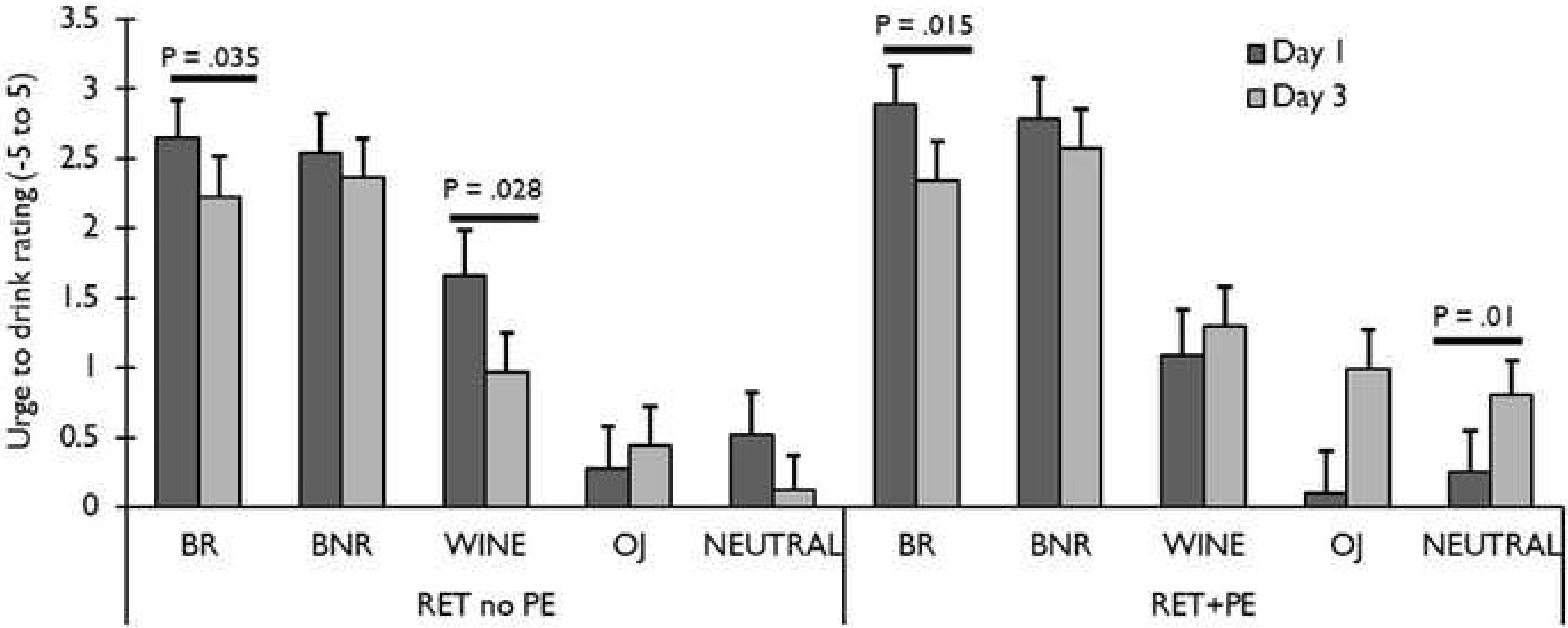


Figure 2 Greyscale

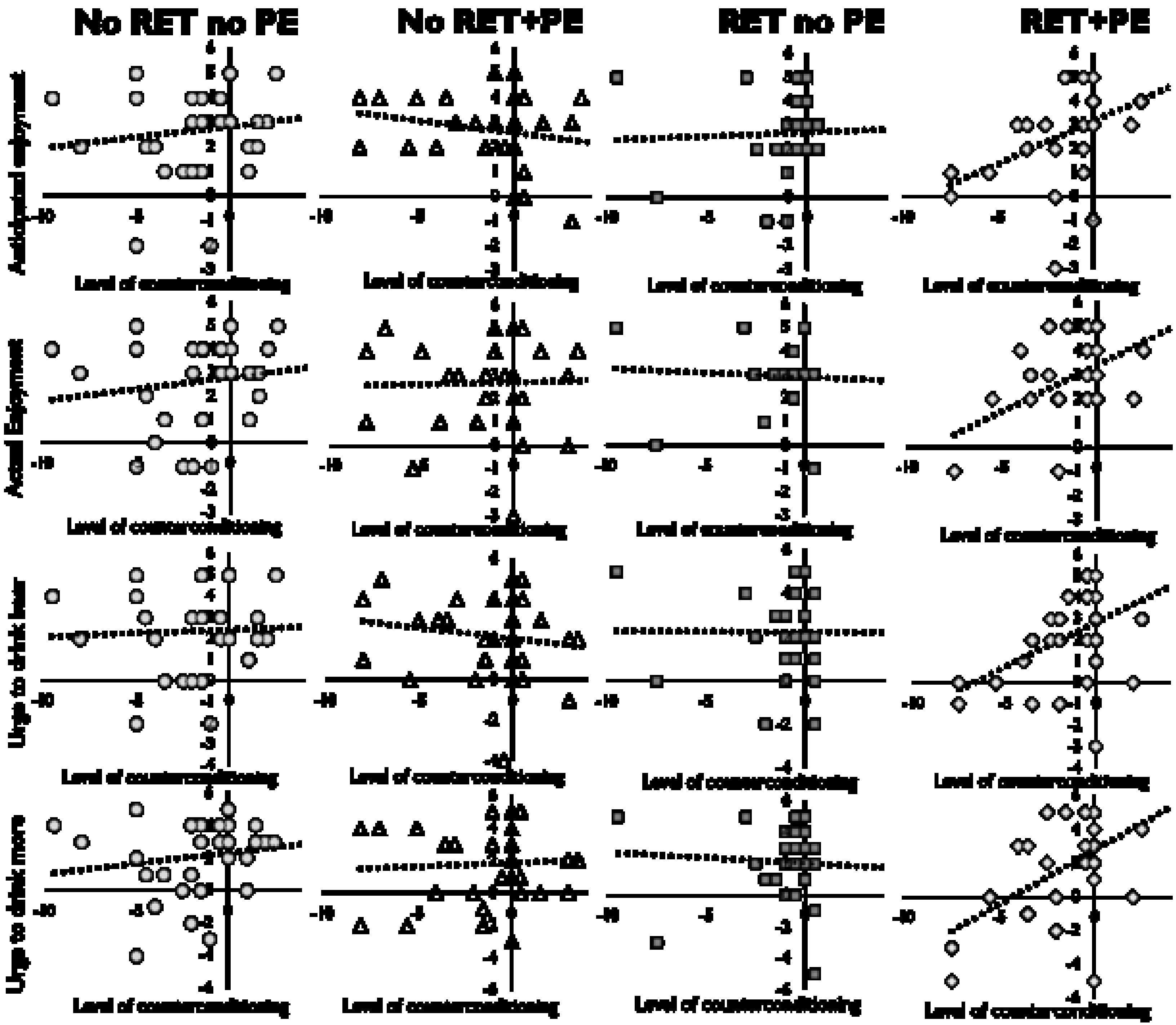


Table 1: Baseline demographic drinking and questionnaire measures. Groups did not differ at false-discovery rate (FDR)-corrected alpha for any variables at baseline. Degrees of freedom for one-way ANOVA are all 3, 116, with the exception of AUDIT data where DFs were 1,83 due to data loss.

		No RET no PE	**No RET + PE**	**RET no PE**	**RET+PE**	**F**	***p***	**FDR adjusted p**
Age		26.13±8.37	27.8±8.65	28.77±11.41	3.07±11.18	.82	.483	>.999
Gender(M:F)		19:11	20:10	21:9	22:8	NA	.59	>.999
AUDIT	Total	18.91±5.03	18.9±4.27	18.23±6.19	18.85±4.27	.094	.963	>.999
	Consumption	8.68±1.17	9±1.21	8.55±1.37	8.15±.93	1.764	.161	>.999
ACQ	COMP	1.9±.9	1.81±.47	1.93±.71	1.87±.66	.165	.920	>.999
	XPECT	3.26±.79	3.16±.79	3.36±.91	3.47±.96	.712	.547	>.999
	PURP	5.4±.9	5.38±.71	5.33±.74	5.43±.81	.084	.969	.995
	EMOT	3±1.08	2.67±1.08	2.83±1.05	2.94±1.21	.529	.663	>.999
	GEN	3.43±.68	3.29±.54	3.37±.59	3.41±.69	.332	.802	>.999
Daily Drinking	Beer (568ml)	2.14±1.33	1.9±1.54	2.22±1.27	1.87±1.52	.457	.713	>.999
	wine (175ml)	.72±1.01	.91±.97	1.02±.92	.92±.84	.538	.657	>.999
	Spirits (25ml)	.94±1.65	1.38±2.39	.8±.74	.91±1.1	.777	.509	>.999
	UK Units (8g EtOH)	8.26±3.86	8.55±4.29	8.68±2.77	9.27±3.72	.395	.757	.991
OCDS	Obsessive	3.77 ± 2.74	3.97 ± 2.89	3.97 ± 3.38	3.4 ± 2.63	.25	.861	>.999
	Compulsive	8.3 ± 2.31	9.27 ± 2.1	9.4 ± 2.34	8.93 ± 2.38	1.39	.251	>.999
CEOA	Sociability	26.3 ± 3.71	25.57 ± 5.32	25.41 ± 3.39	25.97 ± 4.81	.25	.863	.994
	Tension Reduction	7.33 ± 1.86	6.73 ± 2.07	7.66 ± 1.56	7.4 ± 2.27	1.18	.322	>.999
	Liquid Courage	13.03 ± 2.93	12.1 ± 3.03	12.17 ± 3.4	12.7 ± 3.23	.59	.621	>.999
	Sexuality	8.8 ± 2.57	8 ± 2.48	8.34 ± 2.84	8.6 ± 2.99	.48	.696	>.999
	Impairment	18.8 ± 3.42	18.53 ± 4.29	18.41 ± 6.24	18.37 ± 4.14	.05	.984	.984
	Risk Aggression	11.07 ± 3.12	1.83 ± 3.4	1.66 ± 3.07	11.8 ± 3.5	.70	.554	>.999
	Self-perception	6.5 ± 2.16	6.5 ± 2.58	6.38 ± 2.92	5.8 ± 1.94	.57	.635	>.999
SOCRATES	Recognition	17.83 ± 5.41	18.8 ± 6.07	18.8 ± 5.84	15.63 ± 4.37	2.24	.087	>.999
	Ambivalence	12.53 ± 2.96	12.8 ± 3.46	12.13 ± 3.67	11.1 ± 3.48	1.44	.233	>.999
	Taking steps	24.03 ± 6.01	24.27 ± 6.33	22.47 ± 5.95	21.2 ± 6.53	1.61	.191	>.999
BIS/BAS	DRIVE	11.97 ± 2.22	12.03 ± 2.16	11.5 ± 2.5	11.4 ± 2.9	.51	.675	>.999
	FUN	13.5 ± 1.48	14.13 ± 1.5	12.4 ± 2.33	13.97 ± 1.88	5.45	.002	.076
	REWARD	16.8 ± 1.94	17.07 ± 2.03	16.23 ± 2.69	16.63 ± 2.16	.74	.530	>.999
	BIS	2.67 ± 2.89	21.33 ± 2.88	2.13 ± 3.01	2.27 ± 3.04	1.00	.397	>.999
DTS	Tolerance	2.89 ± 1.06	3.11 ± 1.19	2.94 ± 1.1	3.2 ± 1.02	.52	.667	>.999
	Absorption	2.91 ± 1.27	3.13 ± 1.2	3 ± 1.15	3.32 ± 1.14	.67	.570	>.999
	Appraisal	3.24 ± .87	3.38 ± .95	3.26 ± .88	3.43 ± .97	.30	.828	>.999
	Regulation	2.92 ± .96	2.91 ± .92	2.97 ± .98	3.17 ± .93	.48	.700	>.999
	STAI TOTAL	4.23±1.06	39.67 ± 9.26	42.43 ± 11.09	4.83 ± 9.67	.42	.736	.999
STAI	PA TOTAL	34.8 ± 5.92	34.37 ± 5.99	31.37 ± 7.5	36.3 ± 5.89	3.17	.027	.513
PANAS	NA TOTAL	19.03 ± 6.97	18.73 ± 6.1	19.7 ± 7.16	19.2 ± 6.24	.11	.953	>.999
	BDI total	11.83 ± 8.81	1.27 ± 6.6	11.67 ± 9.03	9.4 ± 7.19	.64	.592	>.999

Table 2: Key inferential statistics for cue liking data during the counterconditioning task. Higher-order effects are given in bold, with the simple-effects analyses used to unpick interactions beneath. Beer-Bit CSs= beer cues paired with Bitrex. *Beer-Pic CSs* = Beer cues paired with disgust images, *Neut-Neut CSs* = Neutral images paired with neutral images (control). Superscript letters refer to the terms discussed in the text.

Effect		**ANOVA statistics**	**Text reference**
Trial*Cue Type **interaction**		F(4.134, 475.445)=13.656, *p*<.001, η_p^2=.106	a
Trial Simple effects	*Beer-Bit CSs*	F(3,113)=19.433, *p* <.001, η_p^2=.34	b
	Beer-Pic CSs	F(3,113)=11.274, *p*<.001, η_p^2=.23	c
	Neut-Neut CSs	F(3,113)=0.722, *p*=.512, η_p^2=.02	d
Cue Type*Trial*Retrieval **interaction**		F(4.134,475.445)=2.413, *p*=.046, η_p^2=.021	e
Trial 1 RET > No RET	*Beer-Bit CSs*	F(1,115)=6.936, *p*=.01, η_p^2=.057	f
	Neut-Neut CSs	F(1, 115)=4.594, *p*=.034, η_p^2=.038	g
Trial 1 RET groups	*Cue Type* simple effect	F(1,114)=1.591, *p*=.208, η_p^2=.027	h
Trial 1 No RET groups	*Beer-Pic CSs > Neut-Neut CSs*	F(1,114)=9.353, *p* <.001, η_p^2=.141	i
*Retrieval*Cue Type* interaction	*Trial 4*	F(2,116) =1.867, *p*=.159, η_p^2=.031	j

*Table 3: Reactivity to in-vivo beer: Highest-order (four-way) interaction terms in Day*Retrieval*Responsiveness*PE mixed ANOVAs on anticipated and actual enjoyment of sampled beer and pre and post-drink urge to drink beer. Significant effects are highlighted in bold. Degrees of freedom (DFs)=29 for all t-tests.*

DV	Term DF	F	Sig.	η_p^2	interpretation	Slope in *RET+PE* (*Day 3* score | Responsiveness)
Anticipated enjoyment	4,112	3.416	**.011**	.109	*Day 3* level predicted by counter-conditioning responsiveness only in *RET + PE*	*b*=.355, t=2.56, *p*=.016, η_p^2=.19
Urge to Drink	4,112	5.902	**.007**	.118		*b*=.36, t=2.6, *p*=.015, η_p^2=.194
Actual Enjoyment	4,112	2.321	.061	.077		*b*=.384, t=2.24, *p*=.033, η_p^2=.152
Urge to drink more	4,112	3.048	**.02**	.098		*b*=.641, t=3.1 *p*=.004, η_p^2=.265

Table S1: N respondents in each group at each time point from baseline to final follow up for all drinking-related measures.

			post-	**2**	**3**	**6**	**9**
		baseline	**manipulation**	**weeks**	**months**	**months**	**months**
AUDIT	No RET no PE	22	30	27	25	25	26
	No RET+PE	20	29	24	23	23	26
	RET no PE	22	30	27	23	23	23
	RET+PE	20	29	30	27	27	23
TLFB	No RET no PE	30	30	27	23	23	26
	No RET+PE	30	30	24	21	22	25
	RET no PE	30	30	27	23	23	23
	RET+PE	30	30	29	26	26	23
SOCRATES	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22
ACQ	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22

Table S2: Variability in responses to CSs and UCSs during counterconditioning. Response heterogeneity in 'level' of counterconditioning is evident in the range of liking ratings and standard deviation (SD).

	Min	**Max**	**Mean**	**SD**
Beer-Pic CS liking Trial 1	2.5	10	7.82	1.82
Beer-Pic CS liking Last Trial	0	10	6.51	3.07
Beer-Bit CS liking Trial 1	2.5	10	7.28	1.68
Beer-Bit CS liking Last Trial	0	10	5.83	2.71
Neut-Neut CS liking Trial 1	2.5	10	7.1	1.71
Neut-Neut CS liking last Trial	0	10	7.14	2.39
Δ *Beer-Pic CS liking*	-9.5	4.5	-1.3	2.67
Δ *Beer-Bit CS liking*	-9	3	-1.44	2.47
Δ *Neut-Neut CS liking*	-9.5	5.5	0.05	2.25
Bitrex UCS liking Trial 1	0	6	1.58	1.57
Bitrex UCS liking Last Trial	0	6.5	1.25	1.63
Pic UCS liking Trial 1	0	10	1.58	1.85
Pic UCS liking Last Trial	0	10	1.61	1.89
Neut UCS liking Trial 1	0	10	5.23	2.04
Neut UCS liking Last Trial	0	10	5.76	2.15

Table S3: Pearson's correlations between acute changes in liking of beer cues counter conditioned with Bitrex (*Beer-Bit*) and pictorial (Beer-Pic) UCSs with Day 3 cue and alcohol reactivity outcomes.

		No RET No PE		**No RET + PE**		**RET no PE**		**RET+PE**	
		Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC
Cue image Ratings	**Beer-React liking**	.089	-.101	-.188	-.106	-.113	.21	.178	.212
	Beer-Non-React liking	-.086	-.121	-.1	-.025	-.084	.161	.185	**.37***
	Wine Liking	.085	.141	-.023	-.02	-.075	.278	.275	.322
	OJ Liking	.249	-.131	-.149	-.20	-.096	-.117	-.204	-.071
	Beer-React urge	.043	.073	-.242	-.161	.249	-.013	.171	.295
	Beer-Non-React urge	-.087	-.142	-.154	-.099	.192	-.012	.222	.319
	Wine Urge	.08	.184	-.093	-.026	-.101	-.315	.134	**.39***
	OJ Urge	.093	-.155	-.177	-.083	.225	-.177	.077	**.38***
In vivo beer ratings	**Drink itself liking**	-.016	.066	-.21	-.173	-.154	.101	.324	.058
	Drink itself urge	.037	.086	-.314	-.221	-.262	.155	**.363***	**.441***
	anticipated enjoyment	.137	.074	-.243	-.188	-.35	.052	.247	**.436***
	Anticipatory urge	.046	.046	-.314	-.159	-.349	-.006	.31	**.445***
	Drink enjoyment	.148	.209	-.026	.027	-.143	-.073	.305	**.39***
	Post-drink want more	.161	.098	-.041	.053	-.016	-.075	**.367***	**.515***

SUPPLEMENTARY MATERIALS

METHODS

Counterconditioning trial information:

On each trial, a 'cue image' (CS) was presented alone for 10 second on the left side of the screen in a 400x400 pixel square. This was followed the 'outcome' unconditioned stimulus (UCS). Two negatively reinforcing UCSs were used. The first was 15ml of a 0.067% aqueous solution of denatonium benzoate (Bitrex). This is an extremely bitter solution that reliably produces disgust responses. The second UCS type consisted of four images rated highly for disgust, sourced from the IAPS database. Two of the beer images used as CSs were designated '*Beer Bit CSs*' and would be paired with the Bitrex UCS four times each. The remaining two beer images were designated *Beer Pic CSs* and paired once each with each of the four disgust-induction images from the IAPS database. The designation of beer images to as *Beer Pic* or *Beer-Bit CSs* was random. To control for non-associative effects, two soft drink images were designated '*neutral*' cues and paired with affectively neutral images of office furniture taken from the IAPS database. As both CSs and outcomes in these trials were neutral, they were designated '*Neut-Neut CSs'*.

On *Beer-Bitrex CS* trials, this was a screen saying '*Drink Now'*, prompting consumption of the Bitrex UCS. Eight Bitrex (Bit) UCSs were delivered in total in opaque paper cups. Participants were required to drink all of the liquid in the cup before moving on to the next trial. The remaining number of cups of the Bitrex UCS was unknown by the participant, with the cups themselves stored behind a screen. On *Beer-Pic CS* and *Neut-Neut CS* trials, this was the disgusting or neutral UCS image displayed for 10 seconds, as appropriate. On each trial, the CS image appeared for ten seconds during which time participants participants rated the CS's pleasantness. The 'outcome' UCS then appeared for another ten seconds while participants either looked at the outcome image (*Beer-Pic* and *Neut Neut CS trials*) or drank the Bitrex solution. All images then disappeared and a rating scale for the UCS's pleasantness appeared. All pleasantness ratings were on a scale from 0 (extremely unpleasant) to 10 (extremely pleasant). Counterconditioning was 24 trials in total, consisting of 8 *Beer-Bitrex CS* trials, 8 *Beer-Pic CS* trials and 8 *Neut-Neut CS* trials. Trial types were presented in a pseudo-randomised order with the constraints that no more than two of each type of CS could appear for more than two trials consecutively. Following counterconditioning, all participants were given a square of milk chocolate to mitigate the taste of Bitrex.

Statistical Approach and data handling

Statistical Approach:

Data analysis was performed using IBM SPSS 25 for Windows. Where sphericity was violated in repeated measures, the Greenhouse Geisser correction or multivariate terms were used, depending on ε values and according to published recommendations[60]. This is reflected in non-integer DFs in reported ANOVAs. Changes in short-term drinking-related dependent variables (measure in-lab) were assessed with 2 x 2 x 2mixed ANOVA: within-subjects factor = *Day* (pre-manipulation vs. post-manipulation), between-subjects factors = *Retrieval* (RET vs No RET) and PE (PE, no PE). For analysis of the counterconditioning task, factors of *Cue Type* (Beer-Bit CS/ Beer-Pic CS/ Neut-Neut CS) and *Trial* (1st, 2nd, 3rd, final) were included. The four levels of the *Trial* factor were calculated by taking the mean ratings of each two consecutive presentation of each *CS Type*. Significant interactions in omnibus were investigated with multivariate simple effects analyses and paired tests on marginal means, where appropriate.

Long-term drinking levels were (mean daily beer consumption, mean daily UK units) were analysed using linear mixed models with fixed factors of *Retrieval* and *PE* across *Time* (6:Baseline, Post-manipulation, 2 weeks, 3 months, 6 months, 9 months), modelling per-participant random intercepts as baseline values. *Time* slopes were initially modelled as fixed, with all factorial interactions then allowed to vary randomly, assessing improvement in model fit according to reduction in Bayesian information criterion (BIC) and chi-square tests on -2 log likelihood (-2LL). A reduction >2 in BIC represents an improvement in complexity-penalised model fit. Mixed models were estimated using maximum likelihood with unstructured working correlation matrices. Due to the presence of a small number of unfeasibly high, outlying mean weekly beer consumption values (> 60 units per day, > 400 units/week), analyses were performed on upper-trimmed means with the trim point set at/above 30 units/day. This successfully removed the outlying values from the 2-week time-point, leaving other values unchanged. Rating data during counterconditioning were lost for one participant due to technical error. Alpha for all *a priori* tests was set at 0.05, with *p*-values Bonferroni-corrected for post-hoc tests. For tests of baseline trait, drinking and demographics difference, the False Discovery Rate (FDR) correction was applied [61]All tests are 2-sided. Data were analysed fully blind to condition.

Response attrition at follow-up

Attrition in response was seen at each in all groups at each follow-up time-point. *Table S1,* below gives the respondent Ns at each time point split by group.

Table S1: N respondents in each group at each time point from baseline to final follow up for all drinking-related measures.

		baseline	post-manipulation	2 weeks	3 months	6 months	9 months
AUDIT	No RET no PE	22	30	27	25	25	26
	No RET+PE	20	29	24	23	23	26
	RET no PE	22	30	27	23	23	23
	RET+PE	20	29	30	27	27	23
TLFB	No RET no PE	30	30	27	23	23	26
	No RET+PE	30	30	24	21	22	25
	RET no PE	30	30	27	23	23	23
	RET+PE	30	30	29	26	26	23
SOCRATES	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22
ACQ	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22

RESULTS:

Manipulation Checks:

Variability in learning across the counterconditioning task as well as responses to the UCSs themselves was evident across the sample. Some participants showed very large reductions in *Beer Bit* and *Beer Pic CS* liking, while others showed *increases* in liking of these CSs across the task, despite clear pairing with aversive UCSs. Equally, while most participants rated the *Pic* and *Bitrex* as highly unpleasant, some rated the pictures as 'extremely pleasant' and some even rated the Bitrex above the median point on the scale (i.e. slightly pleasant). Central and dispersion statistics for these ratings are given in *Table S1*. Unlike responses to disgust picture-paired beer images, the change in liking of Bitrex-paired images did not exhibit strong predictive effects on subsequent reactivity to alcohol cues and beer. This is in line with lower variance in response to the Bitrex-paired images during counterconditioning and to Bitrex itself. With rare exceptions, consumption of Bitrex evokes a more potent aversive response than the 'disgust pictures', which may partly explain why the predictive power of 'counterconditioning responsiveness' is lower over long-term follow up.

Table S2: Variability in responses to CSs and UCSs during counterconditioning. Response heterogeneity in 'level' of counterconditioning is evident in the range of liking ratings and standard deviation (SD).

	Min	**Max**	**Mean**	**SD**
Beer-Pic CS liking Trial 1	2.5	10	7.82	1.82
Beer-Pic CS liking Last Trial	0	10	6.51	3.07
Beer-Bit CS liking Trial 1	2.5	10	7.28	1.68
Beer-Bit CS liking Last Trial	0	10	5.83	2.71
Neut-Neut CS liking Trial 1	2.5	10	7.1	1.71
Neut-Neut CS liking last Trial	0	10	7.14	2.39
∆ *Beer-Pic CS liking*	-9.5	4.5	-1.3	2.67
∆ *Beer-Bit CS liking*	-9	3	-1.44	2.47
∆ *Neut-Neut CS liking*	-9.5	5.5	0.05	2.25
Bitrex UCS liking Trial 1	0	6	1.58	1.57
Bitrex UCS liking Last Trial	0	6.5	1.25	1.63
Pic UCS liking Trial 1	0	10	1.58	1.85
Pic UCS liking Last Trial	0	10	1.61	1.89
Neut UCS liking Trial 1	0	10	5.23	2.04
Neut UCS liking Last Trial	0	10	5.76	2.15

Table S3: Pearson's correlations between acute changes in liking of beer cues counter conditioned with Bitrex (Beer-Bit) and pictorial (Beer-Pic) UCSs with Day 3 cue and alcohol reactivity outcomes.

		No RET No PE		**No RET + PE**		**RET no PE**		**RET+PE**	
		Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC
Cue image Ratings	**Beer-React liking**	.089	-.101	-.188	-.106	-.113	.21	.178	.212
	Beer-Non-React liking	-.086	-.121	-.1	-.025	-.084	.161	.185	**.37***
	Wine Liking	.085	.141	-.023	-.02	-.075	.278	.275	.322
	OJ Liking	.249	-.131	-.149	-.20	-.096	-.117	-.204	-.071
	Beer-React urge	.043	.073	-.242	-.161	.249	-.013	.171	.295
	Beer-Non-React urge	-.087	-.142	-.154	-.099	.192	-.012	.222	.319
	Wine Urge	.08	.184	-.093	-.026	-.101	-.315	.134	**.39***
	OJ Urge	.093	-.155	-.177	-.083	.225	-.177	.077	**.38***
In vivo beer ratings	**Drink itself liking**	-.016	.066	-.21	-.173	-.154	.101	.324	.058
	Drink itself urge	.037	.086	-.314	-.221	-.262	.155	**.363***	**.441***
	anticipated enjoyment	.137	.074	-.243	-.188	-.35	.052	.247	**.436***
	Anticipatory urge	.046	.046	-.314	-.159	-.349	-.006	.31	**.445***
	Drink enjoyment	.148	.209	-.026	.027	-.143	-.073	.305	**.39***
	Post-drink want more	.161	.098	-.041	.053	-.016	-.075	**.367***	**.515***

Success of memory reactivation procedures:

Motivational impact of retrieval cues:
'Liking' of relevant drink cues (beer or orange juice) during the retrieval/no retrieval manipulation was assessed with *RET X PE X Cue Type* ANOVA. For this analysis, the liking ratings were averaged for the relevant 'retrieval' images (Beer images in RET groups and orange juice images in No RET groups) and for the 'neutral' drink cues (coffee and cola images in all groups). This revealed a main effect of *Cue Type* and a *Cue Type x Retrieval x PE* interaction [F(1,116) = 5.429, *p* =.024, η_p^2= .043]. Comparison of the simple effects of *Cue Type* within each group showed that the relevant reactivation cues were liked more than the neutral coffee/cola neutral cues in all groups (all F(1, 116) > 5.475, *p*<.021, η_p^2> .045) except for the *No RET + PE* group, where the orange juice images was not significantly greater than the cola/coffee images [F(1, 116) = 3.708, *p* = .057, η_p^2 = .031]. No between-group differences were observed. '*Urge to drink*' the relevant drink (beer in RET groups, or orange juice in No RET) in response to retrieval cues showed main effects of *Cue Type* [F(1, 116) = 123.075, *p*<.0001, η_p^2 = .515] and *Retrieval* [F(1, 116) = 5.703, *p*=.019, η_p^2 = .047]. In all groups, *urge to drink* was higher in response to the relevant retrieval cues than neutral drink (coffee/cola) cues. Cue-induced *urge to drink* beer in the *RET* groups was lower than cue-induced urge to drink orange juice in the *No RET* groups.

Motivational impact of in-vivo drink reward: Pre the prediction-error generation procedure, there were no group differences in liking of (*ps* >.719 η_p^2s<.001) anticipated enjoyment of (*ps* >.685 η_p^2s<.001) or urge to drink (*ps* >.719 η_p^2s<.001) the *in vivo* sample of beer or orange juice. In the *No PE* groups (where the drinks were actually consumed during retrieval) there was no group difference between actual enjoyment of the drinks [F(1,58) = .223, *p*=.639, η_p^2 = .004] nor *desire to drink more* of the drink [F(1,58) = .142, *p*=.708, η_p^2 = .003]. In total, this indicates that the *RET* and *No RET* procedures were well matched in terms of their ability to engage hedonic and motivational consumption processes.

Prediction error generation: Withholding drink reward in PE groups is intended to induce cognitive prediction error or '*surprise*'. Analysis of rated 'surprise' levels following the retrieval and PE/no PE procedures showed a main effect of PE, indicating greater surprise following the PE procedure than the no PE procedure drink [F(1,116) = 309.79, p<.001, η_p^2 = .728]. This did not interact with *Retrieval* group. The PE generation procedure was thus highly successful and equally effective in *RET* and *no RET* groups. In the two *PE* groups, Spearman correlations indicated that larger PE was predicted by greater prior anticipated *liking of the drink* [ρ(60)=-.428,p=.001], greater beer cue- induced *urge to drink* [ρ(60)=-.415, p=.001] and greater *liking of beer cues* [ρ(60)=-.337 p=.008], confirming the intuitive proposition that strength of cognitive PE is a function of anticipation of reward. Invoked anticipation of reward by retrieval cues may explain why previous clinical studies have shown reconsolidation interference effects in the absence of explicit manipulation of PE. Note that the negative sign of the correlation is due to the negative coding of surprise, with -5 being 'extremely unexpected'.

Counterconditioning

***Aversiveness of UCSs*:** A main effect of UCS Type (Bitrex > Disgusting Picture > neutral picture) was observed [F(2,230)=284.791, p<.0001, η_p^2= .712], along with a *UCS Type X Trial* interaction. The interaction indicated cumulative aversion in response to Bitrex UCS, with pleasantness ratings becoming more extremely negative across *Trials* [Trial simple effect for Bitrex F(3,113) = 5.712, p = .001, η_p^2= .132]. There were no effects or interaction with *Retrieval* or *PE* groups. *Overall*, the disgusting UCSs were thus effective negative reinforcers during counterconditioning.

Predictors of response to counterconditioning and changes in drinking.

Disgust propensity and sensitivity were predictive of alcohol consumption during the post-manipulation period, with greater general propensity to disgust [r(120) = -.31, p = .001] and sensitivity to disgusting stimuli [r(120) = -.365, p<.001] predicting lower total alcohol consumption. Disgust propensity was also associated with participants' mean ratings of the unpleasantness of the disgusting images during counterconditioning, indicating higher rated unpleasantness with greater disgust propensity [r(120) = -.311, p = .001]

Pleasantness ratings of the Bitrex UCSs were negatively associated with post-manipulation AUDIT scores [r(117) = -.259 p = .005] and urge to drink in response to beer images [r(119) = -.213 p = .005]. In-lab ratings of reactivity were moderately, (but significantly) correlated with questionnaire-measured craving and drinking outside of the lab (*rs 0.2 – 0.39, p*s 0.01-0.029).

Peri-reactivation affect and arousal may be key moderators of counterconditioning effects, since counterconditioning is an inherently aversive procedure which may interact with negative affect and anxiety in strengthening learning. Further, emotional arousal is well established to potentiate associative learning. Indeed, *arousal* induced by exposure to drug stimuli without reinforcement has been posited as a possible explanation for the enhancing effect of retrieval-extinction procedures, rather than memory rewriting [36].

In support of this interpretation, pre-counterconditioning state anxiety levels on the STAI modestly negatively predicted beer total drinking levels [*r*(120)=-.268, *p*=.003] post-manipulation and at 2-week follow up [*r* (107) = -.214, *p* =.027], but not 3 months, 6 months or 9 months. Similarly, negative affect, derived from the PANAS predicted lower drinking levels at post-manipulation [*r* (120) = -.188, *p* =.04] and 2 weeks [*r* (107) = -.212, *p* =.029] but not longer-term follow ups periods. This is consistent with the engagement of dual processes; affective potentiation of counterconditioning (new learning), yielding shorter-term effects on maladaptive drinking behaviour, with a reconsolidation-based *rewriting* mechanism accounting for more durable long-term reductions in drinking.

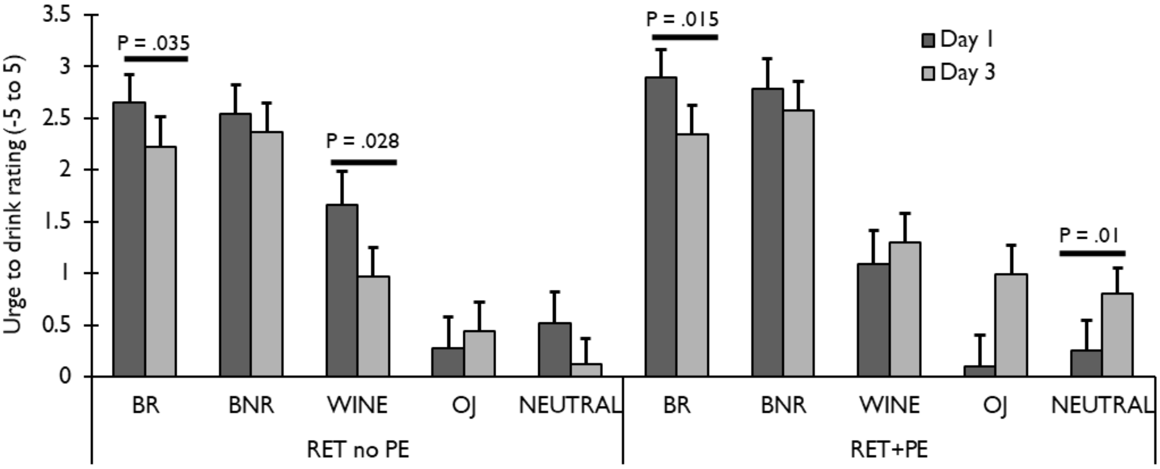
Exploratory subgroup analysis of 'responders': Analysis of only participants who were responsive to counterconditioning (defined as those who reduced their liking of *Beer-Pic* AND *Beer-Bit* cues from the first to last trial of counterconditioning, as is common in conditioning literature), yielded group *N*s of No *RET no PE* =15, *No RET+PE*=10, *RET no PE*=10, *RET+PE*=15. Thus only half, or fewer, of participants acutely displayed 'full' counterconditioning of cues. Re-analysis of reactivity to the beer with *Day X RET X PE* ANOVAs in these groups revealed trend-level *Day*RET*PE* interactions for *urge to drink* [F(1,46)=3.17, *p*=.082, η_p^2= .064]. Multivariate simple effects analyses revealed that this was due to an effect of *Retrieval* in the *PE groups* on *Day 3*, representing lower *urge to drink* in *RET+PE* than *No RET+PE* [F(1,46)=5.281, *p*=.026, η_p^2= .103].

Cue reactivity: Responses to cue images

Ratings of cue image pleasantness and urge to drink depicted beverages during the cue reactivity task were assessed by *Day* (Baseline, post-manipulation) x Retrieval (RET vs No RET) x Prediction Error (PE+/PE-) x Cue Type (Reactivated beer, Non-reactivated beer, wine, orange juice, soft drink) mixed ANOVA, with counterconditioning responsiveness included modelled as a covariate in a fully factorial model.

Urge to drink depicted beverages A *Cue Type* main effect [multivariate F(4,112)=49.353, *p*<.001, η_p^2=.638] and *Day*Cue Type* interaction [multivariate F(4,112)=7.059, *p*<0.001, η_p^2=.201] were found, subsumed under a *Retrieval*PE*Day*Cue Type* interaction [multivariate F(4,113)=3.823, *p* =.006, η_p^2= .12]. The latter interaction was investigated by splitting the analysis by *RET* vs *No RET* groups. A *Day*Cue Type*PE* interaction was present only in the *RET* groups [F(2.424, 191.915)=3.615, *p*=.011, η_p^2= .06]. Inspection of the simple multivariate effects of *Day* indicated that *RET+PE* displayed a significant *reduction* in induced *urge to drink* depicted beer in response to reactivated/ counterconditioned beer cues (*Beer RET*; F(1,57)= 5.5856, *p*=.019, η_p^2=.093) and a significant *increase* in urge to drink orange juice [F(1,57)= 7.293, *p* =.009, η_p^2= .113]. Conversely, in *Ret no PE,* decreases were seen in urge to drink reactivated beer cues [F(1,57)=4.659, *p*=.035, η_p^2=.076] and wine cues F(1,57)=5.771, *p*= .02, η_p^2=.092].

Figure S1: Effects of counterconditioning in *RET* groups on *urge to drink* depicted beverages post-manipulation in the cue reactivity task. BR = reactivated beer images, BNR = Non-reactivated beer images, Wine = Wine images, OJ = Orange juice images, Neutral=coffee/cola images. Bars represent mean±SEM



Liking of cues:

A main effect of *Cue Type* [p<0.001 η_p^2=.164) was found, subsumed under a *Day*Cue Type*Responsiveness* interaction [F(3.351,385.39)=3.905, p=.007, η_p^2=.033]. Analyses on each *Cue Type* showed *Day*Responsiveness* interactions for reactivated beer images [F(1, 115)=4.673, p =.033, η_p^2=.039] and non-reactivated beer images [F(1, 115)=4.665, p=.033, η_p^2=.039], but not wine, orange juice or soft drink (neutral) images. For both types of beer image, greater counterconditioning responsiveness predicted lower *Day 3* liking.

Follow-up data secondary measures

Craving (ACQ-NOW)

General self-rated craving according to the ACQ-NOW did not change in the short-term between pre-and post-manipulation [F(1,116)=1.19, p=.278, η_p^2=.01], however mixed-model analysis with random slopes for *Time* showed long-term reduction in craving over the follow-up period up to 9 months [F(1,106.194)=260.895, p<0.001]. Contrasts on estimated marginal means demonstrated significant reductions in craving by the 2 week follow up [F(1,82)=87.98, p <.001, η_p^2=.518] that persisted or further reduced at all follow-up time points up to at least 9 months [F(1,82)=284.9, p <.001, η_p^2=.777].

Readiness to Change (SOCRATES): All groups reported greater recognition of the need to change their drinking behaviour [F(1,116)=7.378, p=.008, η_p^2=.06], reductions in ambivalence towards their excessive drinking [F(1, 116)=8.897, p=.003, η_p^2=.071] and increases in 'taking steps' to reduce their drinking [F(1, 116)=16.11, p<.001, η_p^2=.122], from baseline to post-manipulation. These beneficial changes did not differ according to *RET* or *PE* group.

Response to Reviewer Comments: PSM-D-19-01615

We would like to thank the reviewers for their considered and constructive comments on our manuscript. They have highlighted some important areas in which we could have been clearer or more concise in our reporting. We have attempted to address all of these issues or otherwise provide clarification in our itemised response below.

Reviewer 1:

General comments: Thank you for the positive appraisal of the manuscript and the useful suggestions for further analysis and presentation of the data, we feel making these amendments has greatly improved the manuscript and its potential utility for readers.

Major Points:

1. *"In figure one, I understand that the data from NoRet +PE and NoRet NoPE are pooled and Ret+ PE and Ret NoPE are pooled as the liking rating is performed before the PE manipulation? It would be informative to be able to see that the prospective PE-manipulated groups within each condition of retrieval were equivalent at that point of the experiment. Particular as the rate of counter conditioning seems to be important in the subsequent analyses"*

The reviewer is correct; In *Figure 1*, the data are pooled across retrieval conditions to represent the interaction reported in the accompanying ANOVA. The interaction indicated that the difference in response to CSs by trial varied across levels of prior 'retrieval' and was not dependent upon 'PE'. The reviewer highlights an important point, however, in that we have not presented the 'liking' and 'wanting' ratings collected as part of the 'retrieval' manipulation. Since this is a subset of the images presented during the 'true' baseline cue reactivity task on Day 1, we considered these of secondary importance to present given the word limit. There were no differences between groups at the point of memory retrieval for any of the measures of cure responsivity (all Fs (1,58) < 2.01, *p*s > 0.15) For completeness, we have now included this this important clarification data. (P13, highlighted).

2. *For clarification it may be worth elaborating on whether the difference between the Retrieval groups is driven by solely by Trial 1 and are they not significantly different at Trial 4? In other words that at the end of counter conditioning both groups reduced liking of beer images to a similar extent.*

Thank you for this useful suggestion. There are no group-level differences by the end of counterconditioning, and we have added the following [P14, highlighted].

Importantly, however, on *Trial 4*, there were no differences across RET conditions in ratings of cues [F(2,116) =1.867, *p*=.159, η_p^2=.031] indicating that absolute responses to counterconditioned cues were similar across groups.

3. *From figure two, the 'level of counterconditiong' for Beer-PIC CS liking is used as it had the greatest variance. For completeness, was there no predictive relationship found for Beer-Bit counterconditioning? This might be interesting to expand on as it seemed to have a greater rate of 'effect' than Beer-PIC in reducing pleasantness (though perhaps this is not significantly different). However, responsiveness to counter conditioning was not predictive of reduction in actual beer consumption, so how to marry these findings together?*

The reviewer is right to note the disparate predictive effects of counterconditioning responsiveness on proximal and longer-term measures and we have further expounded upon this in the discussion, adding the following (P21, highlighted)

Interestingly, the ‘degree’ of counterconditioning was predictive of proximal markers of responding to alcohol, but not long-term drinking outcomes. We believe this is a largely statistical phenomenon, due to greater variance in drinking levels vs. in-lab measures of cue reactivity. However, it is possible that with passing time since reconsolidation-intervention and possible ‘schematisation’ of updated associations, the degree of acute ‘responsiveness’ to counterconditioning becomes less critical to outcomes. This would need to validated empirically, but further highlights a potential disparity between proximal and enduring measures of intervention response and underscores the importance of long-term follow-up.

The change in liking of Bitrex-paired images did not exhibit strong predictive effects on subsequent reactivity to alcohol cues and beer. This is in line with lower variance in response to the Bitrex-paired images during counterconditioning and to Bitrex itself. With rare exceptions, consumption of Bitrex evokes a more potent aversive response than the ‘disgust pictures’, which may partly explain why the predictive power of ‘counterconditioning responsiveness’ is lower over long-term follow up.

We have now added the above discussion to the Supplementary Materials (P3, highlighted).

4. From table S3, I take it that the Ret+PE group had an increased acute liking of in vivo beer ratings, along with the CS images eliciting higher 'urges' (albeit not significantly for all stimuli). Although I note there was not a higher consumption in this group versus the others at the post-manipulation time point, would this increased acute craving not be potentially problematic for further translation?

We would like to clarify that the figures in *Table S3* represent correlations between change scores (i.e. ‘responsiveness’) during counterconditioning and the listed outcomes. These correlations therefore represent standardised metrics of *association* and say nothing about differences in mean level between groups. The *RET+PE* group did not have an increased acute liking or urge vs. the other groups, but it was only in *RET+PE* that liking and urge were *predicted* by counterconditioning responsiveness.

Minor Points:

1. *There are some formatting errors with regards to spacing and repetition of words, also in the references section. In terms of phrasing and expression, in the second methods paragraph on statistics the description of treatment outlying values could be clearer as was the 'upper-trimmed means' used only for the 2-week data? Also, the proportion of males versus females are not given, which could be useful as they were recruited under different limits.*

Apologies for these errors, we have now re-checked the manuscript and amended any errors we were able to find. There were clearly some import errors in Mendeley! We have also clarified the data handling section regarding the trim of outlying values (Methods, page 12 & Supplementary, Page 2). The number of males/females in each group is given in *Table 1*, from which proportions can be easily calculated by readers should they choose.

Did the authors check any differences in when the manipulation was carried out? Whether it was more effective for the 'younger' (48 v 72h) memories? As might be suggested at least from some animal work.

The putative ‘target’ memories were not learned on *Day 1*, but were naturalistic memories that could be years old. Given this, we would not expect the latency between *Day 1* and *Day* 2 to affect the outcomes. Indeed, we have now checked this and confirmed there was no association between latency and primary outcomes. We have added a line at the beginning of the results section (P14) to this effect.

2. *For further discussion should space allow, was the expected enjoyment of the beer related in any way to the extent of prediction error seen? If the heterogeneity of counterconditioning is so important are there any thresholds or criterion that could be offered for future studies from these data as the authors mention in the discussion?*

This is a very interesting and astute suggestion. We have now examined correlations between the pre-prediction error ratings and level of PE achieved. Indeed, as one might expect, there are associations between anticipatory liking and 'urge to drink' the drink and subsequent PE in the two PE groups (No RET+PE and RET+PE), with greater anticipation predicting greater subsequent PE. This supports our assertion in the discussion that high anticipation of reward is likely to produce PE naturally in clinical studies where PE is not specifically manipulated. While this information does not readily lend itself to establishing *absolute* criteria for achieved PE based upon anticipatory liking, it is reasonable to conclude that only those who experience evoked urge-to drink and anticipatory enjoyment from alcohol cues will experience sufficient PE to destabilise memory. This is consistent with our prior work with Nitrous Oxide (Das et al, 2018). We have added the following to the discussion (Page 20):

This is supported by the association between anticipated liking and urge to drink observed and subsequent PE seen in the current study (see Supplementary Materials). In clinical populations, where craving/desire to use is likely to be high to response to drug cues, we may reasonably expect greater PE when drug is not consumed.

And cautiously add the following on page 21:

As a minimum criterion, 'reactivation' cues should evoke an urge/desire to consume and anticipatory enjoyment of drug reward. These measures may be predictive of outcome variability where PE is not assessed.

We have now reported these correlations in the *Supplementary Material* (Page 5) as follows:

In the two *PE* groups, Spearman correlations indicated that larger PE was predicted by greater prior anticipated *liking of the drink* [ρ(60)=-.428,*p*=.001], greater beer cue- induced *urge to drink* [ρ(60)=-.415, *p*=.001] and greater *liking of beer cues* [ρ(60)=-.337 *p*=.008], confirming the intuitive proposition that strength of cognitive PE is a function of anticipation of reward. Invoked anticipation of reward by retrieval cues may explain why previous clinical studies have shown reconsolidation interference effects in the absence of explicit manipulation of PE. Note that the negative sign of the correlation is due to the negative coding of surprise, with -5 being 'extremely unexpected'.

Reviewer #3:

General comments: We thank Reviewer 3 for their kind comments on our manuscript.

1. *The introduction is quite long, it would be better if it could be summarised.*

We have attempted to reduce the length of the introduction as much as possible, removing ~150 words.

2. *Regarding the results section, I found it a bit hard to read due to the amount of data presented. I think it would be better if it could be simplified leaning on the information that can be found in the tables without repeating it. Instead of presenting all the numeric data it would be nicer to read if it was presented in a narrative way with the most relevant numeric results.*

We apologise for the density of statistics presented. Wherever possible, we have attempted to move statistics to tables and describe the results only narratively in the text. For counterconditioning data, (where statistics were most densely presented), we have added a table (Table 2) and all effects are now referenced from the text to this table.

3. *As stated by the authors, the sample size is quite small and large individual variability in the level of achieved counterconditioning was found. Therefore, although this behavioural approach is interesting, more research is still needed.*

We completely agree with the reviewers comment and believe we have highlighted these limitations and need for more research in the manuscript.

4. *In line with the previous comment, there are other aspects regarding the sample selection that may induce some level of bias. For instance, the sample was composed by volunteers, who did not have associated medical comorbidities, and who only consumed beer. Is it expected to conduct future studies in patients with other drinking patterns and different associated medical or psychological problems? This could be stated in the limitation or in future studies section. Also, 50 to 70% of patients diagnosed with alcohol use disorder present some level of cognitive impairment. Could this interfere in an intervention based on memory retrieval and counterconditioning and therefore be an exclusion criteria for receiving the intervention?*

These are important considerations and it will certainly be necessary to replicate these findings in clinically diagnosed samples with AUD, something we plan to pursue in future studies. We selected the current study population due to concerns about potential treatment interference and iatrogenic harm from a relatively untested psychological intervention if conducted in a treatment-seeking group. We would like to clarify that the sample did not *only* drink beer, rather this was their preferred and primary drink. The issue of potential cognitive impairment the author raises is an interesting one and certainly cognitive resistance to adaptive neuroplasticity may be one of the key reasons for lack of response to learning and plasticity-based interventions This study is unable to address this issue, however. We have now added to the discussion highlighting the importance of replication in clinically diagnosed samples and the potential problem of cognitive impairment for this approach (page 23).

5. Describing addiction as a learned pattern of maladaptive alcohol-consumption behaviours could be perceived as a "simplistic" way of describing a disorder with a multi-causal nature, in which biological or social elements (among others) play a role in the maintenance of the addiction. Are the presented techniques expected to be effective on their own, or to be deployed in combination with other therapies? The answer to the question could also be added to the discussion.

It was not our intention to over-simplify addiction’s complex aetiology and apologise if we appear to have to have done so. In the introduction we state:

‘AUDs arise via repeated environmental exposure to alcohol **amid multivariate risk factors** (Sher *et al*. 2005).’

We have now amended the following sentence:

Harmful alcohol consumption may therefore be conceptualised partly a *learned* pattern of maladaptive behaviours (Drummond *et al.* 1990; Hyman 2005). (Page 4, introduction)

We thank the reviewer for the prompt to clarify how we envisage reconsolidation-update mechanisms being incorporated into therapies and feel that this was an important oversight. We have now added the following to the discussion:

Incorporating prior retrieval/destabilisation of MRMs offers a potential means to make these interventions ‘*stick*’, vastly enhancing their long-term efficacy and protecting against relapse. The ‘single-shot’ nature of reconsolidation-interference means it could readily be included as part of a comprehensive psychological treatment program with minimal addition to therapist/patient burden. It may potentially act synergistically with other treatment components that target the biological, cognitive and social causes of AUD by addressing a core, low-level relapsogenic mechanism. (Page 19)

TITLE PAGE:

Long-term Behavioural Rewriting of Maladaptive Drinking Memories via Reconsolidation-Update Mechanisms.

Authors: Gale, Grace., [1] Hennessy, Vanessa E., [1] Walsh, Katie., [1] Kamboj, Sunjeev K. [1] & Das, Ravi. K.[1]*

*Corresponding author. All correspondence to Dr. Ravi Das, Clinical, Educational and Health Psychology, UCL, 26 Bedford Way, London WC1H 0AP, United Kingdom, Email: ravi.das@ucl.ac.uk. Telephone 07341311832

Financial Support: This work was funded by a Medical Research Council grant awarded to Sunjeev Kamboj (grant number: MR/M007006/1)

Conflicts of interest: None

Ethical Standards: *The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013.*

WORD COUNT: 4472

ABSTRACT

Background: Alcohol use disorders can be conceptualised as a learned pattern of maladaptive alcohol-consumption behaviours. The memories encoding these behaviours centrally contribute to long-term excessive alcohol consumption and are a key therapeutic target. The transient period of memory instability sparked during memory reconsolidation offers a therapeutic window to directly *rewrite* these memories using targeted behavioural interventions. However, clinically-relevant demonstrations of the efficacy of this approach are few. We examined key retrieval parameters for destabilising naturalistic drinking memories and the ability of subsequent counterconditioning to effect long-term reductions in drinking.

Methods: Hazardous/harmful beer-drinking volunteers (N=120) were factorially randomised to retrieve (RET) or not retrieve (No RET) alcohol reward memories with (PE) or without (No PE) alcohol reward prediction error. All participants subsequently underwent disgust-based *counterconditioning* of drinking cues. Acute responses to alcohol were assessed pre-and post-manipulation and drinking levels assessed up to 9 months.

Results: Greater long-term reductions in drinking were found when counterconditioning was conducted following retrieval (with and without PE), despite a lack of short-term group differences in motivational responding to acute alcohol. Large variability in acute levels of learning during counterconditioning were noted. 'Responsiveness' to counterconditioning predicted subsequent responses to acute alcohol in *RET+PE* only, consistent with reconsolidation-update mechanisms.

Conclusions: The longevity of behavioural interventions designed to reduce problematic drinking levels may be enhanced by leveraging reconsolidation-update mechanisms to rewrite

maladaptive memory. However, inter-individual variability in levels of corrective learning is likely to determine the efficacy of reconsolidation-updating interventions and should be considered when designing and assessing interventions.

INTRODUCTION

Harmful drinking and alcohol use disorders (AUDs) represent leading causes of global preventable mortality, contributing to 3 million deaths annually (WHO Global status report on alcohol and health, 2018) and recent research suggesting an alarming increase in the prevalence of problem drinking in some demographic groups (Grant et al., 2017). Extant treatments for AUD enjoy limited long-term efficacy, with under 20% completing treatment free of dependence and fewer still maintaining abstinence long-term (Public Health England, 2018). Treatment approaches targeting the fundamental processes underlying the development and maintenance of harmful drinking are required to address this global health priority.

AUDs arise via repeated environmental exposure to alcohol amid multivariate risk factors (Sher et al., 2005). Harmful alcohol consumption may therefore be conceptualised partly a *learned* pattern of maladaptive behaviours (Drummond et al., 1990; Hyman, 2005). Alcohol, like other addictive drugs, induces plasticity in mesocorticolimbic motivational circuitry (Pierce & Kumaresan, 2006). This system supports reward learning, adapting behaviour to seek and maximise rewards when environmental cues signal their availability. Alcohol can therefore support behavioural adaptation towards hyper-motivated alcohol seeking and consumption in the presence of environmental 'trigger' cues. Practically, this manifests as arousal, and a strong desire to drink (craving) in response to certain alcohol-predictive contexts and stimuli (e.g. the sight or smell of beer) (Self, 1998; Sinha & Li, 2007).

Memories that support a harmful level of alcohol use, by linking environmental cues to alcohol reward can be considered to be '*maladaptive reward memories*' (MRMs). Once formed through repeated naturalistic exposure to alcohol with accruing drinking episodes (Robbins et al., 2008), these MRMs are highly robust and display remarkable persistence (Hyman & Malenka, 2001) even after extended periods of abstinence. They therefore believed to be a core substrate underlying persistent relapse susceptibility.

Their central pathogenic role suggests MRMs should be a primary target in the treatment of AUDs (Tronson & Taylor, 2013). A novel approach for directly and permanently ameliorating the negative influence of MRMs on behaviour is to leverage the process of memory *reconsolidation* (Milton & Everitt, 2012; Torregrossa & Taylor, 2013). This is a retrieval-dependent memory maintenance process that serves to strengthen and/or update consolidated memory traces when new memory-relevant information is presented at retrieval. Such updating necessitates the temporary *destabilisation* of memory traces, such that new information can be incorporated and the relevant adjustments to the dendritic and synaptic architecture encoding the memory trace made (Clem & Huganir, 2010; Merlo et al., 2015). If adaptive learning (for example, extinction) is timed correctly following retrieval/destabilisation, such that it occurs in the critical (~2 hour) 'reconsolidation window' when memories are active and unstable, it is theoretically possible to *rewrite* maladaptive memory content to a benign form (Germeroth et al., 2017; Monfils & Holmes, 2018). By re-formatting MRMs such that trigger cues do not provoke alcohol seeking, it may be possible to reduce alcohol consumption and prophylactically guard against relapse over the long-term.

Although a nascent field, there are highly promising early demonstrations of the potential of this approach (Walsh et al., 2018). Extinction (i.e. exposure therapy) following retrieval of MRMs has been shown to produce long-lasting reductions in drug-cue-induced craving and physiological arousal (Xue et al., 2012), and reduce smoking in cigarette smokers (Germeroth et al., 2017). However, there have also been notable failures to replicate reconsolidation-interference effects, particularly using the retrieval-extinction paradigm (Baker et al., 2013; Luyten & Beckers, 2017; Soeter & Kindt, 2011). There are several potential reasons for such discrepant results.

Firstly, extinction itself may represent a sub-optimal 'corrective' learning modality, since it is a largely passive procedure, involving no response from participants, unobserved inter-individual variability in engagement and responsiveness to extinction (Shumake et al., 2018) may mask effects. A promising alternative – *counterconditioning*- re-pairs cues reward cues (e.g. pictures of beer) with negatively-valenced outcomes (e.g. disgust-inducing bitter liquids and images). Disgust- counterconditioning may provide a more potent corrective learning experience than extinction (Tunstall et al., 2012) since it 1) leverages a potent food-rejection mechanism (Rozin & Fallon, 1987) 2) the 'disgust' response to certain images and bitter liquids are powerful and virtually universal (Schienle et al., 2015) and 3) it is an 'active' procedure, meaning participants cannot simply disengage from the task, as occurs during extinction. We have shown broad short-term abolition of attentional biases and reactivity to alcohol cues when *counterconditioning* was conducted after MRM retrieval in hazardous drinkers (Das et al., 2015) a finding that has been further demonstrated in experimental animals (Goltseker et al., 2017), however this has never been shown to affect long-term drinking outcomes.

Secondly, memory retrieval and destabilisation are not synonymous. Indeed, memory destabilisation is highly dependent upon various '*boundary condition*s'(Elsey & Kindt, 2017; Walker & Stickgold, 2016). Primary amongst these are the *length* of retrieval (N cues presented), with retrievals that are either too short or too long failing to spark destabilisation (Merlo et al., 2014, 2018; Suzuki et al., 2004) and the presence of an appropriate 'mismatch' learning signal - *prediction error* (PE)(Schultz et al., 1997; Waelti et al., 2001) - at retrieval (Das et al., 2015; Krawczyk et al., 2017; Sevenster et al., 2013). Specifically, some level of mismatch between predicted and actual outcomes is required for destabilisation (Agustina López et al., 2016; Pedreira et al., 2004).

These key parameters have not been systematically manipulated in clinically-focussed reconsolidation interference studies (Walsh et al., 2018). It is unsurprising, then, that findings are inconsistent. In order to properly assess whether rewriting of alcohol MRMs can be reliably achieved through purely behavioural reconsolidation manipulations, systematic investigation of the role of MRM retrieval and prediction error prior to corrective learning is required.

In the current study, we addressed this issue by systematically manipulating MRM retrieval and the presence of prediction error at retrieval prior to a counterconditioning intervention in heavy drinkers. We assess whether the effects of counterconditioning on cue reactivity and drinking levels are potentiated in a retrieval and prediction error-dependent manner, consistent with reconsolidation-based memory rewriting.

METHODS:

Participants & design: 120 hazardous, beer-preferring drinkers were randomised in a 2 (MRM retrieval/ no retrieval) x 2 (prediction error/ no prediction error) factorial design. All participants completed three sessions, corresponding to *baseline* (on Day 1), retrieval/counterconditioning *manipulation* (Day 3-5) and *post-manipulation* (Day 10 – 13). Primary inclusion criteria were : Ages 18-60 , scoring >8 on the Alcohol Use Disorders Identification Test (AUDIT)(Saunders et al., 1993); Consuming > 40 (men) or >30 (women)

UK units/week (1 unit=8g ethanol), drinking ≥4 days each week, primarily drinking beer, and having non-treatment seeking status. Exclusion criteria were: Pregnancy/breastfeeding, diagnosis of AUD/SUDs, current diagnosed psychiatric disorder, AUD as defined by the SCID; use of psychoactive medications, use of illicit drugs > 2x /month.

Measures:

Questionnaire assessments: The comprehensive effects of alcohol questionnaire (CEOA ;Fromme et al., 1993) retrospectively assessed responses to alcohol, the AUDIT, obsessive-compulsive drinking scale (OCDS; Anton et al., 1995) and alcohol craving questionnaire (ACQ-NOW; Singleton et al., 1994) measured maladaptive drinking patterns. Motivation to reduce drinking was measured by the stages of change readiness and treatment eagerness scale (SOCRATES; Miller & Tonigan, 1996). Distress tolerance and sensitivity to disgust were assessed by the Distress Tolerance Scale (DTS; Simons & Gaher, 2005) and Disgust Propensity and Sensitivity Scale (DPSS-R; Olatunji et al., 2007), respectively. Changes in anxiety and affect due to the counterconditioning procedure were assessed using the state version of the Spielberger State-Trait Anxiety Inventory (STAI-S; Spielberger, 2010) and positive and negative affect scale (PANAS; Watson et al., 1988), respectively. Drinking was quantified using the Timeline Follow-Back diary procedure (Sobell & Sobell, 1992). Depressive symptomatology was assessed with the Beck Depression Inventory (BDI)(Beck et al., 1988).

Cue reactivity assessment: As in our previous study (Das et al., 2019), participants were presented with a 150ml glass of beer and told they would consume this after rating a series of images. They then rated their *urge to drink* and *liking of* four ‘orange juice cue’ images and four ‘beer cue’ images. These were subsequently used as retrieval cues in the ‘no retrieval’

(‘No RET’) and retrieval (‘RET’) procedures respectively on the *manipulation* day. Three *wine* and two soft drink (*neutral*) images (not used as retrieval cues) were also rated, followed by *urge to drink* the *in vivo* beer and *predicted enjoyment* of the beer. These were all rated on 11-point (0 to 10) scales. Participants then consumed the beer according to timed on-screen prompts and rated their post-consumption *actual enjoyment* of the beer and *urge to drink more* beer. These scales thus assessed the acute hedonic and motivational properties of alcohol. These *baseline* (*Day 1*) procedures both allowed assessment of changes in cue reactivity and reinforcing properties of alcohol, and set the expectation of beer consumption to maximise PE on the *manipulation* day when the drink was unexpectedly withheld in PE groups during the appropriate retrieval procedure.

MRM retrieval/PE procedure was one we have previously used to reactivate alcohol MRMs and is described fully elsewhere(Das et al., 2015; Das et al., 2019) . Participants’ MRMs were retrieved by viewing/rating beer cues (*RET*). Control memories were retrieved by viewing/rating orange juice cues (*No RET*). This was identical to the cue reactivity task except 1) the *in vivo* beer was replaced with orange juice in the *No RET* groups 2) only four condition-appropriate cue images were rated. To manipulate prediction error (*PE*), the drink given to participants (orange juice or beer) was unexpectedly withheld by an on-screen prompt reading ‘*Stop, do not drink!*’ in *PE* groups: (*RET+PE* and *No RET+PE)* generating negative prediction error. In the ‘*no PE*’ conditions (*RET no PE, No RET no PE*), the drink was consumed as on *Day 1,* as expected.

Counterconditioning: All four groups underwent counterconditioning after the retrieval/PE manipulations as previously described(Das et al., 2018). Briefly, after a 5-minute interval during which participants completed high working memory load distractor tasks (digit span, prose recall), they were shown four beer images and two neutral drink images (coffee and cola) four times each in a pseudo-randomised, fixed order. Two of the beer images (nominated '*Beer-Bit CSs'*) were paired with consumption of 15ml of a highly bitter solution (.067% aqueous Denatonium Benzoate/*Bitrex*). The other two beer images (nominated '*Beer-Pic CSs'*) were followed by one of four images taken from the IAPS database rated highly for induction of disgust. Two coffee and cola images (nominated '*Neut-Neut* CSs') were followed by neutral rated images from the IAPS database. All pairings occurred on a 100% reinforcement ratio. Full information is given in the *supplementary materials*.

Procedure:

Participants responding to study advertisements were screened for eligibility by telephone. On *Day 1*, (*baseline*), participants attended UCL and completed informed consent before being breathalysed (Lion 500 Alcometer) to ensure abstinence from alcohol. They then completed demographic information (gender, age, education and smoking status) and questionnaire measures (AUDIT, Timeline follow-back, OCDS, CEOA, SOCRATES, DTS and BDI). Participants then completed the cue reactivity and acute beer rating, as described above and in the *supplementary materials*.

On *Day 2* (*manipulation: Day 1* + 48-72hrs), breath-alcohol verified abstinence was confirmed prior to completion of the DPSS-R, ACQ-NOW, PANAS and STAI. Participants then underwent group-appropriate retrieval/no-retrieval and PE/no PE manipulation followed by counterconditioning. After completion of counterconditioning participants re-completed the PANAS. On *Day 3* (*post-manipulation:* 7±2 days after *Day 2*) participants attended the test centre for the final time and recompleted all baseline questionnaires and cue reactivity/ acute beer challenge before debriefing.

Remote follow-up assessments of perceived drinking changes, TLFB, ACQ-NOW and SOCRATES measures were completed at 2 weeks, 3, 6 and 9 months following *Day 3*. Participants were reimbursed at the standard university hourly rate (£10) for in-lab testing sessions and incentivised with an extra £5 for each completed remote follow-up.

Sample size was calculated in G*Power 3.1.9.2 for 1-β=.95 to detect a minimum effect size of $n_p{}^2$=.05 at α=.05 for the interaction in a mixed ANOVA, assuming *ρ* of .5. This yielded a total required sample size of N=78 (26 per group). Anticipating minimal attrition, we randomized N=30/group.

Statistical Approach:

See *supplementary materials* for full data-handling. Changes in short-term outcomes (measured in-lab) were assessed with 2 [*Day*: *pre-manipulation* vs. *post-manipulation*) x 2 [*Retrieval: RET* vs *No RET*] x 2 [*PE*: *PE* vs No PE,] mixed ANOVA. For analysis of the cue reactivity, a factor of *Cue Type* (Beer-Bit CS/ Beer-Pic CS/ Neut-Neut CS/Orange Juice/Neutral) was also modelled. For counterconditioning in addition to *RET* and *PE* factors, factors of *Cue Type* (Beer-Bit CS/Beer-Pic CS/Neut-Neut CS) and *Trial* (1st, 2nd, 3rd, final) were included. Where sphericity was violated in repeated measures, the Greenhouse Geisser or multivariate ANOVAs were used, depending on ε values and according to published recommendations(Stevens, 2012). This is reflected in multivariate/non-integer DFs.

Long-term drinking data were analysed using linear mixed models with fixed factors of *Retrieval* and *PE* across *Time* (6:Baseline, Post-manipulation, 2 weeks, 3 months, 6 months, 9 months), modelling per-participant intercepts as baseline values. *Time* slopes were initially modelled as fixed then as random, assessing improvement in model fit according to reduction >2 in Bayesian information criterion (BIC). Due to the presence of highly outlying mean daily unit alcohol consumption values at 2 weeks (~60 units/day, >450/week), an upper-trim on values was performed on means with the trim at 30 units/day. This removed the two outlying data points (males) from the 2-week data, but did not affect other data. Rating data were lost for one participant due to technical error. Alpha for all *a priori* tests was set at .05, with *p*-values Sidak- corrected for post-hoc tests. For tests of baseline trait, drinking and demographics variables, the False Discovery Rate (FDR) correction was applied. Data were analysed blind to condition.

RESULTS:

Participants were largely equivalent at baseline on key variables (see *Table 1*). Due to technical error, post-screening baseline AUDIT data were only available for *No RET no PE* N=22, *No RET+PE* N=20, *RET no PE* N=22, *RET+PE* N=20. There were no differences between groups in number of days between study sessions and this was unrelated to outcomes.

[TABLE 1 HERE]

Counterconditioning: Those in the two retrieval groups were statistically similar in all liking or urge to drink ratings in response to the beer cues and drink used to retrieve MRMs prior to counterconditioning [all Fs (1,58) ≤2.05,*ps* ≥.158]. Inferential statistics for counterconditioning data are given in *Table 2* for clarity. A *Trial*Cue Type* interaction[a] emerged, indicating significant reductions in liking of Bitrex-paired beer CSs[b] and disgust picture-paired beer CSs[c] across trials, with no significant reduction in unreinforced neutral pictures[d]. Counterconditioning thus successfully reduced mean-level *Beer CS* liking. While successful counterconditioning was evident in both *Retrieval* groups, a marginal *Cue Type*Trial*Retrieval* interaction[e] indicated greater liking of *Beer-Bit CSs*[f] and *Neut-Neut CSs*[g] in the *RET* groups vs. *No RET* groups on *Trial 1* of counterconditioning (see *Figure 1*). In the *RET* groups, all *Cue Types* were liked equally on *Trial 1*[h], while in the *No RET* groups liking of *Beer-Pic CSs* was greater than *Neut-Neut CSs*[i] . On *Trial 4* of counterconditioning, *Neut-Neut Css* were liked more than both *Beer* CSs in the *No RET* groups (*ps*≤.014) but not in the *RET* groups (*ps .072 to .956).* Unreinforced pre-exposure to CSs during MRM retrieval may have thus affected the speed and level at which these were differentiated and subsequently counterconditioned as discriminative stimuli. Importantly, however, on *Trial 4*, there were no differences across RET conditions in ratings of cues[j] indicating that absolute responses to counterconditioned cues were similar across groups.

[TABLE 2 HERE]

Counterconditioning response heterogeneity: There was substantial inter-individual variation in ratings of disgust UCSs and CSs across counterconditioning. Descriptive statistics for these ratings are given in *Supplementary Table S2*. Since memory rewriting here is predicated upon level of 'corrective learning' (i.e. effective counterconditioning of beer cues), a measure of '*counterconditioning responsiveness*' was computed as change in liking of CSs across counterconditioning (Trial 4–Trial 1). Greatest variability was seen in ratings of *Beer Pic CSs*. *Responsiveness* was therefore calculated as Trial 4–Trial 1 Δ in *Beer-PIC CS* liking) to be assessed as a predictor in mixed modelling of drinking outcomes and as a covariate where it was correlated with the dependent variable in general linear models (reinforcing effects of beer), including an interaction term with *Group* to assess the difference in the covariate slope across groups. Correlations with key *post-manipulation* outcomes and exploratory analyses of trait predictors of counterconditioning responsiveness are given in *Supplementary Materials* (*Table S3*).

Prediction error generation

Analysis of rated 'surprise' levels following the retrieval and PE/no PE procedures showed a main effect of *PE*, indicating greater surprise in *PE* groups than *no PE* groups [F(1,116) = 309.79, p<.001, η_p^2 = .728]. This did not interact with *Retrieval* group. The PE generation procedure was thus highly successful and equally effective in *RET* and *no RET* groups. Full statistics on manipulation checks for MRM retrieval are given in the *Supplementary materials.*

Primary Outcomes:

Cue reactivity: Reinforcing effects of alcohol

All analyses of reinforcing effects of *in vivo* beer were analysed with *Day* (*baseline* vs. post-manipulation) x *Retrieval* (RET vs. No RET) x *PE* (PE vs. No PE) RMANCOVAs, including counterconditioning *Responsiveness* as a covariate that could interact with *RET*PE*. Four-way interactions were found for pre-consumption *anticipated enjoyment* and *urge to drink* beer and post-consumption (primed) *urge to drink more* beer. Commensurate with the bivariate correlations, the 4-way interactions were driven *Day*Responsiveness* interactions in *RET+PE* only, indicating that degree of achieved counterconditioning predicted *post-manipulation* reactivity to *in-vivo* beer only in the 'active' *RET+PE* group. For *actual enjoyment* of beer (post consumption), counterconditioning responsiveness again predicted *post-manipulation* enjoyment only in *RET+* However, the 4-way interaction did not reach significance. These interaction terms and simple slopes are given in *Table 3*. Scatterplots of bivariate associations are given in *Figure 2*. Analysis of ratings of pictorial cues used in the cue reactivity task are given in the *supplementary materials*.

[TABLE 3 HERE]

[FIGURE 2 HERE]

Drinking levels:

Beer

The random intercepts-only effects mixed model revealed a significant main effect of *Time* [F(1,522.74)=39.027, *p*<.001] and a marginally significant *RET*PE*Time* interaction [F(1,522.74)=3.965, *p*=.047]. The *Time* effect represented a reduction in beer consumption across the follow-up period, with a mean reduction of .23 UK pints/day at each time point [*b*=-.232, *t*(521.5)=2.04, *p*<.0005]. The 3-way interaction represented a greater reduction in drinking across *Time* in *RET+PE* than *No RET+PE* [*b*=.146, *t*=2.06, *p*=.0397], with no differences between the other groups. Model-predicted and true values for this effect are shown below in *Figure 3* panels A and B. Modelling random slopes for *Time* did not improve model fit (BIC 2128.485→2128.919) and yielded non-significant variance in slopes (Z=1.138, *p*=.255). *Responsiveness* to counterconditioning was not a significant predictor [F(1,119.495)=.72, *p*=.679] and was detrimental to parsimonious model fit (BIC 2128.485→2134.752).

Total Units

The random intercepts-only model for total unit consumption data (BIC=3748.009) yielded a significant effect of *Time* [F(1,533.775)=25.487, *p*<.001] and *RET*Time* interaction [F(1, 533.775)=4.937, *p*=.027]. Simple contrasts on the *Time* main effect against baseline drinking levels showed no overall change in drinking from baseline to post-manipulation [*b*=-.69, *t*(511.97)=.706, *p*=.48] or 2 weeks [*b*=-1.196, *t*(516.53)=.1.194, *p*=.233], with a marginal reduction by 3 months [*b*=-1.97, *t*(519.482)=1.925, *p*=.055] and significant reductions by 6 months [*b*=-4.66, *t*(519.48)=4.549, *p*<.001] and 9 months [*b*=-3.65, *t*(521.05)=3.431, *p*=.001]. Parameter estimates for the *RET*Time* interaction showed a greater reduction in drinking across *Time* in *RET* than *No RET* groups [*b*=.575, *t*(531.58)=2.192, *p*=.029]. Within-groups, the slope for the reduction in drinking across time was highly significant in the *RET* groups [*b*=-.923, *t*(51.26)=-5.008, *p*<.0005] but non-significant in the *No RET* groups [*b*=-.3, *t*(53.958)=-1.177, *p*=.245].

Significant variance in slopes [Z=2.781, *p*=.005] and improved model fit [Δ-2LL $\chi^2(2)$=-18.004, *p* <.001, BIC 3748.09→3743.262] when allowing slopes for *Time* to vary indicated that a random slopes effect model was appropriate. This reduced the *RET**Time* effect to only a marginally significant level [*b*=.623, *t*(107.023)=1.999, *p*=.049]. Including counterconditioning *Responsiveness* as a covariate yielded a borderline-significant predictive impact in drinking [F(1,119.518)=3.916, *p*=.05], but was detrimental to parsimonious model fit [3743.262 →3749.194], so was not included in the final model. Actual and mean model-predicted values for the *RET**Time* effect in the final model are shown in *Figure 3* panels C&D.

[FIGURE 3 HERE]

DISCUSSION

We examined the potential for putative memory reconsolidation mechanisms to catalyse the efficacy and longevity of an experimental learning-based intervention in ameliorating maladaptive drinking patterns. We found mixed evidence that supported the long-term utility of a reconsolidation-focussed approach, while highlighting large response variability and potential limitations of a homogenous learning manipulation.

We observed a greater reduction in over the 9 months follow-up period when counterconditioning followed the -putatively 'active' *retrieval* (*RET*) *with prediction error* (PE) manipulation. Greater reductions in non-specific, *total* alcohol consumption were seen in both MRM retrieval groups, although this was not PE-dependent. These results are broadly consistent with counterconditioning updating MRMs via reconsolidation mechanisms, producing lasting beneficial changes in drinking behaviour. That lasting effects on drinking levels are observed after a one-off, purely behavioural manipulation is encouraging and extends our previous work on ketamine, suggesting reconsolidation-focussed therapies may have a bright future in the treatment of SUDs.

The current results extend our previous findings with counterconditioning during the reconsolidation window (Das et al., 2015) and pharmacological blockade of alcohol MRM reconsolidation by ketamine (Das et al., 2019) . While we previously demonstrated *RET* and *PE* –dependent beneficial effects of counterconditioning on computerised in-lab markers of MRMs, changes in responses to actual alcohol and long-term reductions in drinking following have not, until now, been shown using a purely behavioural reconsolidation-update manipulation.

Unexpectedly, the beneficial effects observed here were primarily evident only in the longer-term drinking outcomes but not acute in-lab measures of cue reactivity. The reason for this discrepancy is uncertain. One possibility is lack of sensitivity or limited ecological validity of an in-lab acute assessment of the reinforcing effects of alcohol, since anticipated enjoyment and urge to drink have no impact on whether beer is consumed or not during this test. An emergent and more compelling interpretation is that memory rewriting manipulations display their true utility when participants are exposed to naturalistic ‘high-risk’ relapse scenarios following manipulation. Indeed, previous research has also observed lagged improvements in phobic symptomatology (Soeter & Kindt, 2015) and craving reductions and CO levels in smokers (Germeroth et al., 2017) following a reconsolidation intervention. This is in line with protection against renewal, reinstatement and spontaneous recovery conferred by reconsolidation interference in the experimental literature. The follow-up period used here is the longest of which we are aware in the reconsolidation literature and the potential for these lagged effects highlights the importance of assessing the longevity of effects over extended follow-up.

Short-term improvements are typically seen following learning-based interventions such as cue-exposure therapy, but these are not maintained across time and contexts. Indeed, in the current study, all groups largely displayed improvements in maladaptive drinking behaviours from pre–to-post-manipulation. Incorporating prior retrieval/destabilisation of MRMs offers a potential means to make these interventions '*stick*', vastly enhancing their long-term efficacy and protecting against relapse. The 'single-shot' nature of reconsolidation-interference means it could readily be included as part of a comprehensive psychological treatment program with minimal addition to therapist/patient burden. It may potentially act synergistically with other treatment components that target the biological, cognitive and social causes of AUD by addressing a core, low-level relapsogenic mechanism.

The discrepancy between retrieval and prediction-error-dependent effects on beer vs. all alcohol consumption was unexpected. We and others (Agustina López et al., 2016; Das et al., 2015; Exton-McGuinness et al., 2015; Krawczyk et al., 2017; Sevenster et al., 2014) have previously forwarded PE or 'surprise' at retrieval as a necessary condition for destabilisation of consolidated memories. Hypothetically, PE signals insufficient or inaccurate prediction of outcomes currently stored by the memory trace and necessitates memory destabilisation to allow the memory to update and stay 'relevant'. These findings may seem to suggest that PE is of secondary importance in sparking memory destabilisation and reconsolidation. Indeed, most previous experimental (Milton et al., 2008; Monfils & Holmes, 2018; Saitoh et al., 2017) and clinically applied (Germeroth et al., 2017; Xue et al., 2012, 2017) reconsolidation studies reporting positive findings have not explicitly manipulated PE. There are several key points that should be borne in mind which caution against such an interpretation, however.

It is typical in reconsolidation studies to omit the primary reinforcer during cue-driven retrieval. This will generate a variable level of PE to the extent that reinforcement is expected, despite not explicitly aiming to manipulate PE. In clinical populations, where craving/desire to use is likely to be high to response to drug cues, we may reasonably expect greater PE when drug is not consumed. This is supported by the association between anticipated liking and urge to drink observed and subsequent PE seen in the current study (see Supplementary Materials). This may well account for variability in previous findings. In the current study, although not statistically significant, the *RET+PE* group also showed the steepest overall absolute decrease in overall drinking, meaning unintended PE generation in the *RET no PE* group may have limited power to observe PE-dependent effects. Indeed, peri-retrieval '*surprise*' ratings demonstrated some variability in surprise in the *RET no PE* and *RET+PE* groups, indicating that some level of unintended PE was occurring in the former group and some expectancy of deception in the latter. For clinical translation, there is minimal extra burden involved in explicitly generating and assessing PE during MRM retrieval. Indeed, in treatment scenarios (e.g. in detoxified drug-abusing patients) it would be ethically unacceptable to reinforce patients with abused drugs. Moreover, there are no demonstrations of *inferiority* of PE vs. no PE at retrieval in memory destabilisation, thus the most prudent course of action would be to include PE-generation procedures in experimental and translational retrieval protocols going forward and at the very least assess these explicitly. As a minimum criterion, 'reactivation' cues should evoke an urge/desire to consume and anticipatory enjoyment of drug reward. These measures may be predictive of outcome variability where PE is not assessed.

Limitations:

We have previously assumed a relatively homogenous response to the counterconditioning intervention, given that is leverages very basic learning and aversion mechanisms. The large observed variability in level of achieved counterconditioning or 'responsiveness' demonstrate that this assumption is not tenable. Some participants displayed reductions of in liking of negatively reinforced beer stimuli over half the scale range while others showed little or no change and some even displayed *increased* liking over the course of the task. Equally, some participants did not rate the UCSs as particularly aversive, with some even rating them as mildly pleasant. Having extensively piloted the doses of Bitrex used here ourselves, this is puzzling to us, although genetic polymorphisms moderating bitterness perception may play a key role (Duffy & Bartoshuk, 2000). We further found that disgust propensity, sensitivity and distress tolerance predicted counterconditioning responsiveness, yielding potentially useful trait markers of likely treatment response. However, such individual variability to counterconditioning likely obscured potential group-level differences in responses to the acute alcohol challenge. Interestingly, the 'degree' of counterconditioning was predictive of proximal markers of responding to alcohol, but not long-term drinking outcomes. We believe this is a largely statistical phenomenon, due to greater variance in drinking levels vs. in-lab measures of cue reactivity. However, it is possible that with passing time since reconsolidation-intervention and possible 'schematisation' of updated associations, the degree of acute 'responsiveness' to counterconditioning becomes less critical to outcomes. This would needto validated empirically, but further highlights a potential disparity between proximal and enduring measures of intervention response and underscores the importance of long-term follow-up.

One could reasonably anticipate equal (or greater) response variability when using retrieval-extinction (Shumake et al., 2018); a paradigm that has dominated behavioural memory rewriting research. This may partially explain the inconsistencies and difficulties replicating findings with retrieval-extinction interventions (Baker et al., 2013; Chen et al., 2014; Luyten & Beckers, 2017; Soeter & Kindt, 2011), since a failure to extinguish would preclude any potentiating effect of prior memory retrieval. These observations highlight the importance assessing level of corrective learning, conducting learning to a criterion level or identifying potential low-responders within reconsolidation-updating paradigms.

Variability in learning is perhaps a reason to recommend pharmacological memory-weakening over purely behavioural memory updating approaches in certain populations. Drugs' pharmacodynamic profiles are generally not subject to influence by individual cognitive variables like learning rates, boredom and punishment insensitivity and may be a key option where behavioural approaches fail.

There is no way of assessing whether the *RET+PE* truly destabilised alcohol MRMs and engaged reconsolidation mechanisms (or did so to an equal degree) in all individuals in the current study, since memory destabilisation is a behaviourally silent process. This remains the primary impediment to translational/clinical developments within the reconsolidation field, which is in desperate need of validated biomarkers of memory destabilisation. The lack of triangulation between short-term lab measures and longer-term drinking outcomes compounds this issue in the current study. We have, however, now demonstrated group-level sufficiency of the *RET+PE* procedure used improving clinically-relevant outcomes in five studies (Das et al., 2018; Das et al., 2015; Das et al., 2018, 2019; Hon et al., 2016). Along with the apparently durable effects on drinking observed here, this lends support to the notion that reconsolidation mechanisms were engaged in the current study. While non reconsolidation mechanisms may explain shorter-term effects on outcome, the emergence of divergent effects longer-term observed here are in line with reconsolidation-update.

The current study highlights fundamental questions regarding the parameters that conspire retrieval conspire to determine the fate of memories at retrieval. The future of memory-rewriting interventions will rely upon better understanding of these parameters and individual optimisation of memory destabilisation procedures based therein. Nevertheless, the results obtained here are should energise future research in the field, particularly to assess whether similar effects can be replicated in clinically diagnosed samples where comorbidities and cognitive impairment from chronic alcohol abuse may further complicate implementation.

TABLE AND FIGURE LEGENDS:

Table 1: Baseline demographic drinking and questionnaire measures. Groups did not differ at false-discovery rate (FDR)-corrected alpha for any variables at baseline. Degrees of freedom for one-way ANOVA are all 3, 116, with the exception of AUDIT data where DFs were 1,83 due to data loss.

Table 2: Key inferential statistics for cue liking data during the counterconditioning task. Higher-order effects are given in bold, with the simple-effects analyses used to unpick interactions beneath. Beer-Bit CSs= beer cues paired with Bitrex. *Beer-Pic CSs* = Beer cues paired with disgust images, *Neut-Neut CSs* = Neutral images paired with neutral images (control). Superscript letters refer to the terms discussed in the text.

Table 3: Reactivity to in-vivo beer: Highest-order (four-way) interaction terms in Day*Retrieval*Responsiveness*PE mixed ANOVAs on anticipated and actual enjoyment of sampled beer and pre and post-drink urge to drink beer. Significant effects are highlighted in bold. Degrees of freedom (DFs)=29 for all t-tests.

Figure 1: Liking ratings for the conditioned stimuli across the counterconditioning task. Significant reductions in liking of the Bitrex-paired beer CS (*Beer-Bit CS*) and disgusting image-paired beer CS (*Beer-Pic CS*) were seen in reactivated and non-reactivated groups. However only in *No RET* did the liking of CSs differ on Trial 1. *****=*Beer-Pic>Neut-Neut,* **¥**=*Beer-Pic>Beer-Bit*, **†**=*Neut-Neut>Beer-Bit,* **#**=*Neut-Neut>Beer-Pic.*

Figure 2: Associations between 'strength' of counterconditioning (change in liking of counterconditioned beer cues) anticipated enjoyment, urge to drink, actual enjoyment and urge to drink more beer on the Day 3 beer reactivity test. The correlations were significant only in RET+PE (rightmost column). Dashed lines are ordinary least-squares linear best fit lines.

Figure 3: **Panel A** (top left) changes in mean daily beer consumption (in UK pints) across the study time points in each group. **Panel B (top right)** Mixed model fit values for beer consumption data. A marginally significant *Time*RET*PE* interaction indicated a steeper reduction across Time in *RET+PE* than *No RET+PE* (p=.037). **Panel C***:* Changes in mean daily unit alcohol consumption across the study time points in each group. **Panel D***:* Model fit values for overall alcohol consumption (total UK unit) data. A significant *RET*Time* interaction indicated significant reductions across time in RET groups but not No RET groups. Panels A&C, error bars represent SD. Panels B and D, error bars represent model SEMs.

REFERENCES

Agustina López, M., Jimena Santos, M., Cortasa, S., Fernández, R. S., Carbó Tano, M., & Pedreira, M. E. (2016). Different dimensions of the prediction error as a decisive factor for the triggering of the reconsolidation process. *Neurobiology of Learning and Memory, 136*, 210–219. https://doi.org/10.1016/j.nlm.2016.10.016

Anton, R. F., Moak, D. H., & Latham, P. (1995). The Obsessive Compulsive Drinking Scale: A Self-Rated Instrument for the Quantification of Thoughts about Alcohol and Drinking Behavior. *Alcoholism: Clinical and Experimental Research*, *19*(1), 92–99. https://doi.org/10.1111/j.1530-0277.1995.tb01475.x

Baker, K. D., McNally, G. P., & Richardson, R. (2013). Memory retrieval before or after extinction reduces recovery of fear in adolescent rats. *Learning and Memory*. https://doi.org/10.1101/lm.031989.113

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*(1), 77–100.

Chen, S., Cai, D., Pearce, K., Sun, P. Y. W., Roberts, A. C., & Glanzman, D. L. (2014). Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in Aplysia. *ELife*, *3*, e03896. https://doi.org/10.7554/eLife.03896

Clem, R. L., & Huganir, R. L. (2010). Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. *Science, 330*(6007), 1108–1112.

Das, R. K., Gale, G., Walsh, K., Hennessy, V. E., Iskandar, G., Mordecai, L. A., Brandner, B., Kindt, M., Curran, H. V., & Kamboj, S. K. (2019). Ketamine can reduce harmful drinking by pharmacologically rewriting drinking memories. *Nature Communications*, *10*(1), 5187. https://doi.org/10.1038/s41467-019-13162-w

Das, R.K., Lawn, W., & Kamboj, S. K. (2015). Rewriting the valuation and salience of alcohol-related stimuli via memory reconsolidation. *Translational Psychiatry*, *5*(9), e645–e645. https://doi.org/10.1038/tp.2015.132

Das, R.K., Walsh, K., Hannaford, J., Lazzarino, A. I., & Kamboj, S. K. (2018). Nitrous oxide may interfere with the reconsolidation of drinking memories in hazardous drinkers in a prediction-error-dependent manner. *European Neuropsychopharmacology*, *28*(7), 828–840. https://doi.org/10.1016/j.euroneuro.2018.05.001

Das, R.K., Gale, G., Hennessy, V., & Kamboj, S. K. (2018). A Prediction Error-driven Retrieval Procedure for Destabilizing and Rewriting Maladaptive Reward Memories in Hazardous Drinkers. *Journal of Visualized Experiments*, *56097*(131), e56097–e56097. https://doi.org/10.3791/56097

Drummond, D. C., Cooper, T., & Glautier, S. P. (1990). Conditioned learning in alcohol dependence: implications for cue exposure treatment. *British Journal of Addiction*, *85*(6), 725–743.

Duffy, V. B., & Bartoshuk, L. M. (2000). Food acceptance and genetic variation in taste. *Journal of the American Dietetic Association*, *100*(6), 647–655. https://doi.org/10.1016/S0002-8223(00)00191-7

Elsey, J. W. B., & Kindt, M. (2017). Breaking boundaries: optimizing reconsolidation-based interventions for strong and old memories. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *24*(9), 472–479. https://doi.org/10.1101/lm.044156.116

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories-The role of prediction errors in memory reconsolidation. In *Behavioural Brain Research* (Vol. 278). https://doi.org/10.1016/j.bbr.2014.10.011

Fromme, K., Stroot, E. A., & Kaplan, D. (1993). Comprehensive effects of alcohol: Development and psychometric assessment of a new expectancy questionnaire. *Psychological Assessment*, *5*(1), 19–26. https://doi.org/10.1037/1040-3590.5.1.19

Germeroth, L. J., Carpenter, M. J., Baker, N. L., Froeliger, B., LaRowe, S. D., & Saladin, M. E. (2017). Effect of a Brief Memory Updating Intervention on Smoking Behavior. *JAMA Psychiatry*, *74*(3), 214. https://doi.org/10.1001/jamapsychiatry.2016.3148

Goltseker, K., Bolotin, L., & Barak, S. (2017). Counterconditioning During Reconsolidation Prevents Relapse of Cocaine Memories. *Neuropsychopharmacology*, *42*(3), 716–726. https://doi.org/10.1038/npp.2016.140

Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., Huang, B., Jung, J., Zhang, H., Fan, A., & Hasin, D. S. (2017). Prevalence of 12-Month Alcohol Use, High-Risk Drinking, and DSM-IV Alcohol Use Disorder in the United States, 2001-2002 to 2012-2013. *JAMA Psychiatry*, *74*(9), 911. https://doi.org/10.1001/jamapsychiatry.2017.2161

Hon, T., Das, R. K. R. K., & Kamboj, S. K. S. K. S. K. (2016). The effects of cognitive reappraisal following retrieval-procedures designed to destabilize alcohol memories in high-risk drinkers. *Psychopharmacology*, *233*(5), 851–861. https://doi.org/10.1007/s00213-015-4164-y

Hyman, S. E. (2005). Addiction: a disease of learning and memory. *American Journal of Psychiatry*, *162*(8), 1414–1422.

Hyman, S. E., & Malenka, R. C. (2001). Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature Reviews Neuroscience*, *2*(10), 695–703.

Krawczyk, M. C., Fernández, R. S., Pedreira, M. E., & Boccia, M. M. (2017). Toward a better understanding on the role of prediction error on memory processes: From bench to clinic. *Neurobiology of Learning and Memory*, *142*(Part A), 13–20. https://doi.org/10.1016/j.nlm.2016.12.011

Luyten, L., & Beckers, T. (2017). A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. *Neurobiology of Learning and Memory*. https://doi.org/10.1016/j.nlm.2017.07.014

Merlo, E., Bekinschtein, P., Jonkman, S., & Medina, J. H. (2015). Molecular Mechanisms of Memory Consolidation, Reconsolidation, and Persistence. *Neural Plasticity*, *2015*, 1–2. https://doi.org/10.1155/2015/687175

Merlo, E., Milton, A. L., & Everitt, B. J. (2018). A Novel Retrieval-Dependent Memory Process Revealed by the Arrest of ERK1/2 Activation in the Basolateral Amygdala. *The Journal of Neuroscience*, *38*(13), 3199–3207. https://doi.org/10.1523/JNEUROSCI.3273-17.2018

Merlo, E., Milton, A. L., Goozee, Z. Y., Theobald, D. E., & Everitt, B. J. (2014). Reconsolidation and Extinction Are Dissociable and Mutually Exclusive Processes: Behavioral and Molecular Evidence. *Journal of Neuroscience*, *34*(7), 2422–2431. https://doi.org/10.1523/JNEUROSCI.4001-13.2014

Miller, W. R., & Tonigan, J. S. (1996). Assessing drinkers' motivation for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*, *10*(2), 81–89. https://doi.org/10.1037/0893-164X.10.2.81

Milton, A. L., & Everitt, B. J. (2012). The persistence of maladaptive memory: addiction, drug memories and anti-relapse treatments. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1119–1139. https://doi.org/10.1016/j.neubiorev.2012.01.002

Milton, A. L., Lee, J. L. C., Butler, V. J., Gardner, R., & Everitt, B. J. (2008). Intra-amygdala and systemic antagonism of NMDA receptors prevents the reconsolidation of drug-associated memory and impairs subsequently both novel and previously acquired drug-seeking behaviors. *The Journal of Neuroscience*, *28*(33), 8230–8237.

Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*, *5*(12), 1032–1042. https://doi.org/http://dx.doi.org/10.1016/S2215-

0366%2818%2930270-0

Olatunji, B. O., Cisler, J. M., Deacon, B. J., Connolly, K., & Lohr, J. M. (2007). The Disgust Propensity and Sensitivity Scale-Revised: Psychometric properties and specificity in relation to anxiety disorder symptoms. *Journal of Anxiety Disorders*, *21*(7), 918–930. https://doi.org/10.1016/j.janxdis.2006.12.005

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory*, *11*(5), 579–585.

Pierce, R. C., & Kumaresan, V. (2006). The mesolimbic dopamine system: The final common pathway for the reinforcing effect of drugs of abuse? *Neuroscience & Biobehavioral Reviews*, *30*(2), 215–238. https://doi.org/http://dx.doi.org/10.1016/j.neubiorev.2005.04.016

Public HealthEngland, Department of Health, & National Drug Evidence Centre. (2018). *Adult Drug Statistics from the National Drug Treatment Monitoring System (NDTMS). April 2017*, 38. www.facebook.com/PublicHealthEngland

Robbins, T. W., Ersche, K. D., & Everitt, B. J. (2008). Drug addiction and the memory systems of the brain. *Annals of the New York Academy of Sciences*, *1141*(1), 1–21.

Rozin, P., & Fallon, A. E. (1987). A Perspective on Disgust. *Psychological Review*. https://doi.org/10.1037/0033-295X.94.1.23

Saitoh, A., Akagi, K., Oka, J.-I., & Yamada, M. (2017). Post-reexposure administration of d-cycloserine facilitates reconsolidation of contextual conditioned fear memory in rats. *Journal of Neural Transmission*, *124*(5), 583–587. https://doi.org/10.1007/s00702-017-1704-0

Saunders, J. B., Aasland, O. G., Babor, T. F., Delafuente, J. R., Grant, M., De La Fuente, J. R., Grant, M., Delafuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schienle, A., Arendasy, M., & Schwab, D. (2015). Disgust Responses to Bitter Compounds: the Role of Disgust Sensitivity. *Chemosensory Perception*, *8*(4), 167–173. https://doi.org/10.1007/s12078-015-9186-7

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Self, D. W. (1998). Neural substrates of drug craving and relapse in drug addiction. *Annals of Medicine*, *30*(4), 379–389. https://doi.org/10.3109/07853899809029938

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science*, *339*(6121), 830–833. https://doi.org/10.1126/science.1231357

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory*, *21*(11), 580–584. https://doi.org/10.1101/lm.035493.114

Sher, K. J., Grekin, E. R., & Williams, N. A. (2005). The Development of Alcohol Use Disorders. *Annual Review of Clinical Psychology*, *1*(1), 493–523. https://doi.org/10.1146/annurev.clinpsy.1.102803.144107

Shumake, J., Jones, C., Auchter, A., & Monfils, M.-H. (2018). Data-driven criteria to assess fear remission and phenotypic variability of extinction in rats. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1742), 20170035. https://doi.org/10.1098/rstb.2017.0035

Simons, J. S., & Gaher, R. M. (2005). The Distress Tolerance Scale: Development and Validation of a Self-Report Measure. *Motivation and Emotion*, *29*(2), 83–102.

https://doi.org/10.1007/s11031-005-7955-3

Singleton, E. G., Henningfield, J. E., & Tiffany, S. T. (1994). Alcohol craving questionnaire: ACQ-Now: background and administration manual. *Baltimore: NIDA Addiction Research Centre*.

Sinha, R., & Li, C. S. R. (2007). Imaging stress- and cue-induced drug and alcohol craving: association with relapse and clinical implications. *Drug and Alcohol Review*, *26*(1), 25–31. https://doi.org/10.1080/09595230601036960

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back. In *Measuring alcohol consumption* (pp. 41–72). Springer.

Soeter, M., & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. *Learning & Memory*, *18*(6), 357–366. https://doi.org/10.1101/lm.2148511

Soeter, M., & Kindt, M. (2015). An abrupt transformation of phobic behavior after a post-retrieval amnesic agent. *Biological Psychiatry*, *78*(12), 880–886. https://doi.org/10.1016/j.biopsych.2015.04.006

Spielberger, C. D. (2010). *State‐ Trait Anxiety Inventory*. Wiley Online Library.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *The Journal of Neuroscience*, *24*(20), 4787–4795.

Torregrossa, M. M., & Taylor, J. R. (2013). Learning to forget: manipulating extinction and reconsolidation processes to treat addiction. *Psychopharmacology*, *226*(4), 659–672.

Tronson, N. C., & Taylor, J. R. (2013). Addiction: A drug-induced disorder of memory reconsolidation. In *Current Opinion in Neurobiology* (Vol. 23, Issue 4, pp. 573–580). Elsevier Current Trends. https://doi.org/10.1016/j.conb.2013.01.022

Tunstall, B. J., Verendeev, A., & Kearns, D. N. (2012). A comparison of therapies for the treatment of drug cues: Counterconditioning vs. extinction in male rats. *Experimental and Clinical Psychopharmacology*, *20*(6), 447–453. https://doi.org/10.1037/a0030593

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48.

Walker, M. P., & Stickgold, R. (2016). Understanding the boundary conditions of memory reconsolidation. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 113, Issue 28, pp. E3991–E3992). https://doi.org/10.1073/pnas.1607964113

Walsh, K. H., Das, R. K., Saladin, M. E., & Kamboj, S. K. (2018). Modulation of naturalistic maladaptive memories using behavioural and pharmacological reconsolidation-interfering strategies: a systematic review and meta-analysis of clinical and ‘sub-clinical’ studies. *Psychopharmacology*, *235*(9), 2507–2527. https://doi.org/10.1007/s00213-018-4983-8

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

WHO | Global status report on alcohol and health. (2018). *WHO*. https://www.who.int/substance_abuse/publications/global_alcohol_report/en/

Xue, Y.-X., Deng, J.-H., Chen, Y.-Y., Zhang, L.-B., Wu, P., Huang, G.-D., Luo, Y.-X., Bao, Y.-P., Wang, Y.-M., Shaham, Y., Shi, J., & Lu, L. (2017). Effect of selective inhibition of reactivated nicotine-associated memories with propranolol on nicotine craving. *JAMA Psychiatry*, *74*(3), 224–232. https://doi.org/10.1001/jamapsychiatry.2016.3907

Xue, Y.-X., Luo, Y.-X., Wu, P., Shi, H.-S. H.-S., Xue, L.-F., Chen, C., Zhu, W.-L., Ding, Z.-B., Bao, Y., Shi, J., Epstein, D. H., Shaham, Y., & Lu, L. (2012). A Memory Retrieval-

Extinction Procedure to Prevent Drug Craving and Relapse. *Science*, *336*(6078), 241–245. https://doi.org/10.1126/science.1215070

SUPPLEMENTARY MATERIALS

METHODS

Counterconditioning trial information:

On each trial, a 'cue image' (CS) was presented alone for 10 second on the left side of the screen in a 400x400 pixel square. This was followed the 'outcome' unconditioned stimulus (UCS). Two negatively reinforcing UCSs were used. The first was 15ml of a 0.067% aqueous solution of denatonium benzoate (Bitrex). This is an extremely bitter solution that reliably produces disgust responses. The second UCS type consisted of four images rated highly for disgust, sourced from the IAPS database. Two of the beer images used as CSs were designated '*Beer Bit CSs*' and would be paired with the Bitrex UCS four times each. The remaining two beer images were designated *Beer Pic CSs* and paired once each with each of the four disgust-induction images from the IAPS database. The designation of beer images to as *Beer Pic* or *Beer-Bit CSs* was random. To control for non-associative effects, two soft drink images were designated '*neutral*' cues and paired with affectively neutral images of office furniture taken from the IAPS database. As both CSs and outcomes in these trials were neutral, they were designated '*Neut-Neut CSs'*.

On *Beer-Bitrex CS* trials, this was a screen saying '*Drink Now'*, prompting consumption of the Bitrex UCS. Eight Bitrex (Bit) UCSs were delivered in total in opaque paper cups. Participants were required to drink all of the liquid in the cup before moving on to the next trial. The remaining number of cups of the Bitrex UCS was unknown by the participant, with the cups themselves stored behind a screen. On *Beer-Pic CS* and *Neut-Neut CS* trials, this was the disgusting or neutral UCS image displayed for 10 seconds, as appropriate. On each trial, the CS image appeared for ten seconds during which time participants participants rated the CS's pleasantness. The 'outcome' UCS then appeared for another ten seconds while participants either looked at the outcome image (*Beer-Pic* and *Neut Neut CS trials*) or drank the Bitrex solution. All images then disappeared and a rating scale for the UCS's pleasantness appeared. All pleasantness ratings were on a scale from 0 (extremely unpleasant) to 10 (extremely pleasant). Counterconditioning was 24 trials in total, consisting of 8 *Beer-Bitrex C*S trials, 8 *Beer-Pic CS* trials and 8 *Neut-Neut CS* trials. Trial types were presented in a pseudo-randomised order with the constraints that no more than two of each type of CS could appear for more than two trials consecutively. Following counterconditioning, all participants were given a square of milk chocolate to mitigate the taste of Bitrex.

Statistical Approach and data handling

Statistical Approach:

Data analysis was performed using IBM SPSS 25 for Windows. Where sphericity was violated in repeated measures, the Greenhouse Geisser correction or multivariate terms were used, depending on ε values and according to published recommendations[60]. This is reflected in non-integer DFs in reported ANOVAs. Changes in short-term drinking-related dependent variables (measure in-lab) were assessed with 2 x 2 x 2mixed ANOVA: within-subjects factor = *Day* (pre-manipulation vs. post-manipulation), between-subjects factors = *Retrieval* (RET vs No RET) and PE (PE, no PE). For analysis of the counterconditioning task, factors of *Cue Type* (Beer-Bit CS/ Beer-Pic CS/ Neut-Neut CS) and *Trial* (1st, 2nd, 3rd, final) were included. The four levels of the *Trial* factor were calculated by taking the mean ratings of each two consecutive presentation of each *CS Type*. Significant interactions in omnibus were investigated with multivariate simple effects analyses and paired tests on marginal means, where appropriate.

Long-term drinking levels were (mean daily beer consumption, mean daily UK units) were analysed using linear mixed models with fixed factors of *Retrieval* and *PE* across *Time* (6:Baseline, Post-manipulation, 2 weeks, 3 months, 6 months, 9 months), modelling per-participant random intercepts as baseline values. *Time* slopes were initially modelled as fixed, with all factorial interactions then allowed to vary randomly, assessing improvement in model fit according to reduction in Bayesian information criterion (BIC) and chi-square tests on -2 log likelihood (-2LL). A reduction >2 in BIC represents an improvement in complexity-penalised model fit. Mixed models were estimated using maximum likelihood with unstructured working correlation matrices. Due to the presence of a small number of unfeasibly high, outlying mean weekly beer consumption values (> 60 units per day, > 400 units/week), analyses were performed on upper-trimmed means with the trim point set at/above 30 units/day. This successfully removed the outlying values from the 2-week time-point, leaving other values unchanged. Rating data during counterconditioning were lost for one participant due to technical error. Alpha for all *a priori* tests was set at 0.05, with *p*-values Bonferroni-corrected for post-hoc tests. For tests of baseline trait, drinking and demographics difference, the False Discovery Rate (FDR) correction was applied [61]All tests are 2-sided. Data were analysed fully blind to condition.

Response attrition at follow-up

Attrition in response was seen at each in all groups at each follow-up time-point. *Table S1*, below gives the respondent Ns at each time point split by group.

Table S1: N respondents in each group at each time point from baseline to final follow up for all drinking-related measures.

		baseline	**post-manipulation**	**2 weeks**	**3 months**	**6 months**	**9 months**
AUDIT	No RET no PE	22	30	27	25	25	26
	No RET+PE	20	29	24	23	23	26
	RET no PE	22	30	27	23	23	23
	RET+PE	20	29	30	27	27	23
TLFB	No RET no PE	30	30	27	23	23	26
	No RET+PE	30	30	24	21	22	25
	RET no PE	30	30	27	23	23	23
	RET+PE	30	30	29	26	26	23
SOCRATES	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22
ACQ	No RET no PE	30	30	27	25	26	26
	No RET+PE	30	30	24	23	24	26
	RET no PE	30	30	27	23	22	23
	RET+PE	30	30	30	27	25	22

RESULTS:

Manipulation Checks:

Variability in learning across the counterconditioning task as well as responses to the UCSs themselves was evident across the sample. Some participants showed very large reductions in *Beer Bit* and *Beer Pic CS* liking, while others showed *increases* in liking of these CSs across the task, despite clear pairing with aversive UCSs. Equally, while most participants rated the *Pic* and *Bitrex* as highly unpleasant, some rated the pictures as ‘extremely pleasant’ and some even rated the Bitrex above the median point on the scale (i.e. slightly pleasant). Central and dispersion statistics for these ratings are given in *Table S1*. Unlike responses to disgust picture-paired beer images, the change in liking of Bitrex-paired images did not exhibit strong predictive effects on subsequent reactivity to alcohol cues and beer. This is in line with lower variance in response to the Bitrex-paired images during counterconditioning and to Bitrex itself. With rare exceptions, consumption of Bitrex evokes a more potent aversive response than the ‘disgust pictures’, which may partly explain why the predictive power of ‘counterconditioning responsiveness’ is lower over long-term follow up.

Table S2: Variability in responses to CSs and UCSs during counterconditioning. Response heterogeneity in ‘level’ of counterconditioning is evident in the range of liking ratings and standard deviation (SD).

	Min	**Max**	**Mean**	**SD**
Beer-Pic CS liking Trial 1	2.5	10	7.82	1.82
Beer-Pic CS liking Last Trial	0	10	6.51	3.07
Beer-Bit CS liking Trial 1	2.5	10	7.28	1.68
Beer-Bit CS liking Last Trial	0	10	5.83	2.71
Neut-Neut CS liking Trial 1	2.5	10	7.1	1.71
Neut-Neut CS liking last Trial	0	10	7.14	2.39
∆ *Beer-Pic CS liking*	-9.5	4.5	-1.3	2.67
∆ *Beer-Bit CS liking*	-9	3	-1.44	2.47
∆ *Neut-Neut CS liking*	-9.5	5.5	0.05	2.25
Bitrex UCS liking Trial 1	0	6	1.58	1.57
Bitrex UCS liking Last Trial	0	6.5	1.25	1.63
Pic UCS liking Trial 1	0	10	1.58	1.85
Pic UCS liking Last Trial	0	10	1.61	1.89
Neut UCS liking Trial 1	0	10	5.23	2.04
Neut UCS liking Last Trial	0	10	5.76	2.15

Table S3: Pearson's correlations between acute changes in liking of beer cues counter conditioned with Bitrex (Beer-Bit) and pictorial (Beer-Pic) UCSs with Day 3 cue and alcohol reactivity outcomes.

		No RET No PE		**No RET + PE**		**RET no PE**		**RET+PE**	
		Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC	Δ Beer-BIT	Δ Beer-PIC
Cue image Ratings	**Beer-React liking**	.089	-.101	-.188	-.106	-.113	.21	.178	.212
	Beer-Non-React liking	-.086	-.121	-.1	-.025	-.084	.161	.185	**.37***
	Wine Liking	.085	.141	-.023	-.02	-.075	.278	.275	.322
	OJ Liking	.249	-.131	-.149	-.20	-.096	-.117	-.204	-.071
	Beer-React urge	.043	.073	-.242	-.161	.249	-.013	.171	.295
	Beer-Non-React urge	-.087	-.142	-.154	-.099	.192	-.012	.222	.319
	Wine Urge	.08	.184	-.093	-.026	-.101	-.315	.134	**.39***
	OJ Urge	.093	-.155	-.177	-.083	.225	-.177	.077	**.38***
In vivo beer ratings	**Drink itself liking**	-.016	.066	-.21	-.173	-.154	.101	.324	.058
	Drink itself urge	.037	.086	-.314	-.221	-.262	.155	**.363***	**.441***
	anticipated enjoyment	.137	.074	-.243	-.188	-.35	.052	.247	**.436***
	Anticipatory urge	.046	.046	-.314	-.159	-.349	-.006	.31	**.445***
	Drink enjoyment	.148	.209	-.026	.027	-.143	-.073	.305	**.39***
	Post-drink want more	.161	.098	-.041	.053	-.016	-.075	**.367***	**.515***

Success of memory reactivation procedures:

Motivational impact of retrieval cues:
'Liking' of relevant drink cues (beer or orange juice) during the retrieval/no retrieval manipulation was assessed with *RET X PE X Cue Type* ANOVA. For this analysis, the liking ratings were averaged for the relevant 'retrieval' images (Beer images in RET groups and orange juice images in No RET groups) and for the 'neutral' drink cues (coffee and cola images in all groups). This revealed a main effect of *Cue Type* and a *Cue Type x Retrieval x PE* interaction [F(1,116) = 5.429, *p* =.024, η_p^2= .043]. Comparison of the simple effects of *Cue Type* within each group showed that the relevant reactivation cues were liked more than the neutral coffee/cola neutral cues in all groups (all F(1, 116) > 5.475, *p*<.021, η_p^2> .045) except for the *No RET + PE* group, where the orange juice images was not significantly greater than the cola/coffee images [F(1, 116) = 3.708, *p* = .057, η_p^2 = .031]. No between-group differences were observed. '*Urge to drink*' the relevant drink (beer in RET groups, or orange juice in No RET) in response to retrieval cues showed main effects of *Cue Type* [F(1, 116) = 123.075, *p*<.0001, η_p^2 = .515] and *Retrieval* [F(1, 116) = 5.703, *p*=.019, η_p^2 = .047]. In all groups, *urge to drink* was higher in response to the relevant retrieval cues than neutral drink (coffee/cola) cues. Cue-induced *urge to drink* beer in the *RET* groups was lower than cue-induced urge to drink orange juice in the *No RET* groups.

Motivational impact of in-vivo drink reward: Pre the prediction-error generation procedure, there were no group differences in liking of (*ps* >.719 η_p^2*s*<.001) anticipated enjoyment of (*ps* >.685 η_p^2*s*<.001) or urge to drink (*ps* >.719 η_p^2*s*<.001) the *in vivo* sample of beer or orange juice. In the *No PE* groups (where the drinks were actually consumed during retrieval) there was no group difference between actual enjoyment of the drinks [F(1,58) = .223, *p*=.639, η_p^2 = .004] nor *desire to drink more* of the drink [F(1,58) = .142, *p*=.708, η_p^2 = .003]. In total, this indicates that the *RET* and *No RET* procedures were well matched in terms of their ability to engage hedonic and motivational consumption processes.

Prediction error generation: Withholding drink reward in PE groups is intended to induce cognitive prediction error or '*surprise*'. Analysis of rated 'surprise' levels following the retrieval and PE/no PE procedures showed a main effect of PE, indicating greater surprise following the PE procedure than the no PE procedure drink [F(1,116) = 309.79, p<.001, η_p^2 = .728]. This did not interact with *Retrieval* group. The PE generation procedure was thus highly successful and equally effective in *RET* and *no RET* groups. In the two *PE* groups, Spearman correlations indicated that larger PE was predicted by greater prior anticipated *liking of the drink* [ρ(60)=-.428,p=.001], greater beer cue- induced *urge to drink* [ρ(60)=-.415, p=.001] and greater *liking of beer cues* [ρ(60)=-.337 p=.008], confirming the intuitive proposition that strength of cognitive PE is a function of anticipation of reward. Invoked anticipation of reward by retrieval cues may explain why previous clinical studies have shown reconsolidation interference effects in the absence of explicit manipulation of PE. Note that the negative sign of the correlation is due to the negative coding of surprise, with -5 being 'extremely unexpected'.

Counterconditioning

***Aversiveness of UCSs*:** A main effect of UCS Type (Bitrex > Disgusting Picture > neutral picture) was observed [F(2,230)=284.791, p<.0001, η_p^2= .712], along with a *UCS Type X Trial* interaction. The interaction indicated cumulative aversion in response to Bitrex UCS, with pleasantness ratings becoming more extremely negative across *Trials* [Trial simple effect for Bitrex F(3,113) = 5.712, p = .001, η_p^2= .132]. There were no effects or interaction with *Retrieval* or *PE* groups. *Overall*, the disgusting UCSs were thus effective negative reinforcers during counterconditioning.

Predictors of response to counterconditioning and changes in drinking.

Disgust propensity and sensitivity were predictive of alcohol consumption during the post-manipulation period, with greater general propensity to disgust [r(120) = -.31, p = .001] and sensitivity to disgusting stimuli [r(120) = -.365, p<.001] predicting lower total alcohol consumption. Disgust propensity was also associated with participants' mean ratings of the unpleasantness of the disgusting images during counterconditioning, indicating higher rated unpleasantness with greater disgust propensity [r(120) = -.311, p = .001]

Pleasantness ratings of the Bitrex UCSs were negatively associated with post-manipulation AUDIT scores [r(117) = -.259 p = .005] and urge to drink in response to beer images [r(119) = -.213 p = .005]. In-lab ratings of reactivity were moderately, (but significantly) correlated with questionnaire-measured craving and drinking outside of the lab (*rs 0.2 – 0.39, p*s 0.01-0.029).

Peri-reactivation affect and arousal may be key moderators of counterconditioning effects, since counterconditioning is an inherently aversive procedure which may interact with negative affect and anxiety in strengthening learning. Further, emotional arousal is well established to potentiate associative learning. Indeed, *arousal* induced by exposure to drug stimuli without reinforcement has been posited as a possible explanation for the enhancing effect of retrieval-extinction procedures, rather than memory rewriting [36].

In support of this interpretation, pre-counterconditioning state anxiety levels on the STAI modestly negatively predicted beer total drinking levels [$r(120)=-.268$, $p=.003$] post-manipulation and at 2-week follow up [r (107) = -.214, p =.027], but not 3 months, 6 months or 9 months. Similarly, negative affect, derived from the PANAS predicted lower drinking levels at post-manipulation [r (120) = -.188, p =.04] and 2 weeks [r (107) = -.212, p =.029] but not longer-term follow ups periods. This is consistent with the engagement of dual processes; affective potentiation of counterconditioning (new learning), yielding shorter-term effects on maladaptive drinking behaviour, with a reconsolidation-based *rewriting* mechanism accounting for more durable long-term reductions in drinking.

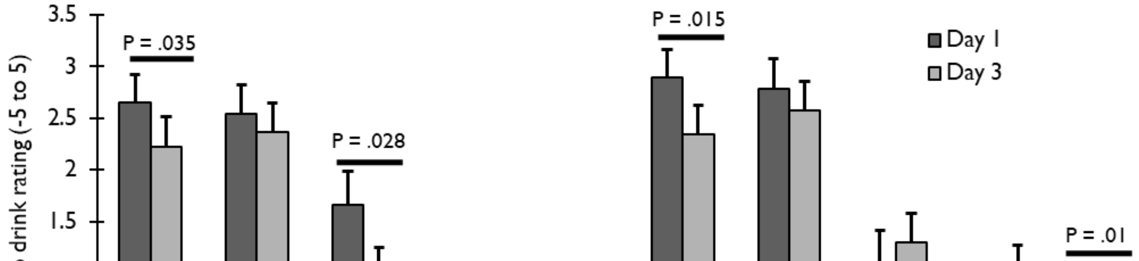
Exploratory subgroup analysis of 'responders': Analysis of only participants who were responsive to counterconditioning (defined as those who reduced their liking of *Beer-Pic* AND *Beer-Bit* cues from the first to last trial of counterconditioning, as is common in conditioning literature), yielded group *N*s of No *RET no PE* =15, *No RET+PE*=10, *RET no PE*=10, *RET+PE*=15. Thus only half, or fewer, of participants acutely displayed 'full' counterconditioning of cues. Re-analysis of reactivity to the beer with *Day X RET X PE* ANOVAs in these groups revealed trend-level *Day*RET*PE* interactions for *urge to drink* [F(1,46)=3.17, p=.082, η_p^2= .064]. Multivariate simple effects analyses revealed that this was due to an effect of *Retrieval* in the *PE groups* on *Day 3*, representing lower *urge to drink* in *RET+PE* than *No RET+PE* [F(1,46)=5.281, p=.026, η_p^2= .103].

Cue reactivity: Responses to cue images

Ratings of cue image pleasantness and urge to drink depicted beverages during the cue reactivity task were assessed by *Day* (Baseline, post-manipulation) x Retrieval (RET vs No RET) x Prediction Error (PE+/PE-) x Cue Type (Reactivated beer, Non-reactivated beer, wine, orange juice, soft drink) mixed ANOVA, with counterconditioning responsiveness included modelled as a covariate in a fully factorial model.

Urge to drink depicted beverages A *Cue Type* main effect [multivariate F(4,112)=49.353, p<.001, η_p^2=.638] and *Day*Cue Type* interaction [multivariate F(4,112)=7.059, p<0.001, η_p^2=.201] were found, subsumed under a *Retrieval*PE*Day*Cue Type* interaction [multivariate F(4,113)=3.823, p =.006, η_p^2= .12]. The latter interaction was investigated by splitting the analysis by *RET* vs *No RET* groups. A *Day*Cue Type*PE* interaction was present only in the *RET* groups [F(2.424, 191.915)=3.615, p=.011, η_p^2= .06]. Inspection of the simple multivariate effects of *Day* indicated that *RET+PE* displayed a significant *reduction* in induced *urge to drink* depicted beer in response to reactivated/ counterconditioned beer cues (*Beer RET*; F(1,57)= 5.5856, p=.019, η_p^2=.093) and a significant *increase* in urge to drink orange juice [F(1,57)= 7.293, p =.009, η_p^2= .113]. Conversely, in *Ret no PE*, decreases were seen in urge to drink reactivated beer cues [F(1,57)=4.659, p=.035, η_p^2=.076] and wine cues F(1,57)=5.771, p= .02, η_p^2=.092].

Figure S1: Effects of counterconditioning in *RET* groups on *urge to drink* depicted beverages post-manipulation in the cue reactivity task. BR = reactivated beer images, BNR = Non-reactivated beer images, Wine = Wine images, OJ = Orange juice images, Neutral=coffee/cola images. Bars represent mean±SEM



Liking of cues:
A main effect of *Cue Type* [p<0.001 η_p^2=.164) was found, subsumed under a *Day**Cue Type*Responsiveness* interaction [F(3.351,385.39)=3.905, p=.007, η_p^2=.033]. Analyses on each *Cue Type* showed *Day**Responsiveness* interactions for reactivated beer images [F(1, 115)=4.673, p =.033, η_p^2=.039] and non-reactivated beer images [F(1, 115)=4.665, p=.033, η_p^2=.039], but not wine, orange juice or soft drink (neutral) images. For both types of beer image, greater counterconditioning responsiveness predicted lower *Day 3* liking.

Follow-up data secondary measures

Craving (ACQ-NOW)
General self-rated craving according to the ACQ-NOW did not change in the short-term between pre-and post-manipulation [F(1,116)=1.19, p=.278, η_p^2=.01], however mixed-model analysis with random slopes for *Time* showed long-term reduction in craving over the follow-up period up to 9 months [F(1,106.194)=260.895, p<0.001]. Contrasts on estimated marginal means demonstrated significant reductions in craving by the 2 week follow up [F(1,82)=87.98, p <.001, η_p^2=.518] that persisted or further reduced at all follow-up time points up to at least 9 months [F(1,82)=284.9, p <.001, η_p^2=.777].

Readiness to Change (SOCRATES): All groups reported greater recognition of the need to change their drinking behaviour [F(1, 116)=7.378, p=.008, η_p^2=.06], reductions in ambivalence towards their excessive drinking [F(1, 116)=8.897, p=.003, η_p^2=.071] and increases in 'taking steps' to reduce their drinking [F(1, 116)=16.11, p<.001, η_p^2=.122], from baseline to post-manipulation. These beneficial changes did not differ according to *RET* or *PE* group.