

## How representative are neuroimaging samples? Large-scale evidence for trait anxiety differences between fMRI and behaviour-only research participants.

Caroline J. Charpentier<sup>1,5\*</sup>, Paul Faulkner<sup>2</sup>, Eva R. Pool<sup>3</sup>, Verena Ly<sup>4</sup>, Marieke S. Tollenaar<sup>4</sup>, Lisa M. Klun<sup>1</sup>, Aniek Fransen<sup>1</sup>, Yumeya Yamamori<sup>5</sup>, Níall Lally<sup>5</sup>, Anahit Mkrtchian<sup>5</sup>, Vincent Valton<sup>5</sup>, Quentin J.M. Huys<sup>5</sup>, Ioannis Sarigiannidis<sup>5</sup>, Kelly A. Morrow<sup>6</sup>, Valentina Krenz<sup>7</sup>, Felix Kalbe<sup>7</sup>, Anna Cremer<sup>7</sup>, Gundula Zerbes<sup>7</sup>, Franziska M. Kausche<sup>7</sup>, Nadine Wanke<sup>7</sup>, Alessio Giarrizzo<sup>3</sup>, Erdem Pulcu<sup>8</sup>, Susannah Murphy<sup>8,9</sup>, Alexander Kaltenboeck<sup>8,10</sup>, Michael Browning<sup>8,9</sup>, Lynn K. Paul<sup>1</sup>, Roshan Cools<sup>11,12</sup>, Karin Roelofs<sup>11</sup>, Luiz Pessoa<sup>6</sup>, Catherine J. Harmer<sup>8,9</sup>, Henry W. Chase<sup>13</sup>, Christian Grillon<sup>14</sup>, Lars Schwabe<sup>7</sup>, Jonathan P. Roiser<sup>5</sup>, Oliver J. Robinson<sup>5</sup>, John P. O'Doherty<sup>1</sup>

<sup>1</sup>California Institute of Technology, Pasadena, CA, USA. <sup>2</sup>University of Roehampton, London, UK. <sup>3</sup>University of Geneva, Geneva, Switzerland. <sup>4</sup>Dept of Clinical Psychology, Leiden University; Leiden Institute for Brain and Cognition, Leiden, the Netherlands. <sup>5</sup>Institute of Cognitive Neuroscience, University College London, London, UK. <sup>6</sup>University of Maryland, College Park, MD, USA. <sup>7</sup>University of Hamburg, Hamburg, Germany. <sup>8</sup>Dept of Psychiatry, University of Oxford, Oxford, UK. <sup>9</sup>Oxford Health NHS Trust, Oxford, UK. <sup>10</sup>Dept of Psychiatry and Psychotherapy, Clinical Division of Social Psychiatry, Medical University of Vienna, Austria. <sup>11</sup>Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. <sup>12</sup>Dept of Psychiatry, Radboud University Medical Centre, Nijmegen, the Netherlands. <sup>13</sup>Dept of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA. <sup>14</sup>National Institute of Mental Health, Bethesda, MD, USA.

\*Corresponding author contact information: [ccharpen@caltech.edu](mailto:ccharpen@caltech.edu)

### Abstract

Over the past three decades, functional MRI (fMRI) has become key to study how cognitive processes are implemented in the human brain. However, the question of whether participants recruited into

fMRI studies differ from participants recruited into other study contexts has received little to no attention. This is particularly pertinent when effects fail to generalize across study contexts: for example, a behavioural effect discovered in a non-imaging context not replicating in a neuroimaging environment. Here, we tested the hypothesis, motivated by preliminary findings (n=272), that fMRI participants differ from behaviour-only participants on one fundamental individual difference variable: trait anxiety. Analysing trait anxiety scores and possible confounding variables from healthy volunteers across multiple institutions (n=3317), we found robust support for lower trait anxiety in fMRI study participants, consistent with a sampling or self-selection bias. The bias was larger in studies that relied on phone screening (compared to full in-person psychiatric screening), recruited at least partly from convenience samples (compared to community samples), and in pharmacology studies. Our findings highlight the need for surveying trait anxiety at recruitment and for appropriate screening procedures or sampling strategies to mitigate this bias.

**Keywords:** trait anxiety, neuroimaging, behaviour, sampling bias

## Introduction

Neuroimaging methods, such as functional MRI (fMRI), have been fundamental to the emergence of cognitive neuroscience as a research field. These methods provide a unique window into the function of the human brain and into the implementation of cognitive processes at the computational, neural and network levels. However, a key question that has not been examined in the field is whether individuals who participate in fMRI studies differ from those who participate in behaviour-only studies in terms of their psychological or psychiatric profiles. Given that many studies in cognitive neuroscience involve a behavioural piloting phase to assess behavioural effects, followed by an fMRI scanning phase to assess neural mechanisms, it is important to ensure that individuals who volunteer to participate in each study context exhibit similar profiles and can be characterized by similar population distributions. This is especially relevant for studies in which effects that are present (and replicate) outside the scanner (Bolton and Robinson, 2017) fail to replicate (Garibbo *et al.*, 2019) inside the scanning environment. Similarly, some effects may be more easily found in fMRI than in behavioural studies, due to higher alertness and/or stress associated with the scanner environment. While there is evidence that physical characteristics of the scanning environment, such as acoustic noise (Hommel *et al.*, 2012; Skouras *et al.*, 2013; Kobald *et al.*, 2016), can affect cognitive and affective processes, their neural basis, and

hormonal responses (Gossett *et al.*, 2018), poor generalizability across testing contexts could also be due, in part, to unanticipated biases in study recruitment.

Specifically, anxiety is likely to be a key factor influencing individuals' decisions to select themselves into specific studies, situations, or environments. Here, we formally test the hypothesis that, because of this selection bias or because of variability in screening procedures, individuals who participate in fMRI studies exhibit lower trait anxiety than individuals who participate in behavioural studies. Within populations of healthy volunteers, it is likely that anxious individuals are more nervous about going into the MRI scanner, and are discouraged or excluded from participating if claustrophobic (Meléndez and McCrank, 1993; Katz *et al.*, 1994; Murphy and Brunberg, 1997). While perhaps not unexpected, the hypothesis of lower trait anxiety in fMRI study contexts has to our knowledge never been formally tested, nor do we know the extent to which the distribution of anxiety levels is likely to be reduced to a narrower range.

In addition, this question is also particularly pertinent for studies in which a modulatory effect of anxiety on behaviour is expected and for researchers interested in the mechanisms of anxiety per se. While anxiety disorders constitute a major global health burden (Beddington *et al.*, 2008), anxiety is also a normative adaptive function that varies across the general population. Studying anxiety in healthy human subjects can thus help bridge the gap between animal models of anxiety and clinical applications for patients with anxiety disorders (Grillon *et al.*, 2019; Robinson *et al.*, 2019). Myriad studies have suggested that a wide range of cognitive functions are modulated by anxiety levels (see (Robinson *et al.*, 2013) for a review): sensory processing and gating (Grillon, 2002; Engel-Yeger and Dunn, 2011; Poli and Angrilli, 2015), attentional biases toward negative emotional stimuli (Bar-Haim *et al.*, 2007; Cisler and Koster, 2010), decreased emotion regulation (Etkin *et al.*, 2010; Farmer and Kashdan, 2012), deficits in attentional control (Bishop, 2009), reduced working memory performance (Shackman *et al.*, 2006; Yao *et al.*, 2018), impairments during reinforcement learning (Browning *et al.*, 2015; Mkrtchian *et al.*, 2017) and increased risk avoidance during decision-making (Maner *et al.*, 2007; Clark *et al.*, 2012; Charpentier *et al.*, 2017). Neuroimaging studies have provided evidence for heightened amygdala responses to negative emotional stimuli (Etkin *et al.*, 2004; Stein *et al.*, 2007)<sup>0</sup> and reduced connectivity between the prefrontal cortex and the amygdala (Etkin *et al.*, 2010; Robinson *et al.*, 2014; Shackman *et al.*, 2016; Carlisi and Robinson, 2018) in anxiety. Because of this multifaceted association between anxiety and cognition, many behavioural and neuroimaging studies in cognitive neuroscience routinely collect measures of anxiety. A common self-report measure of anxiety can be obtained from the State-Trait Anxiety Inventory or STAI (Spielberger *et al.*, 1983). Trait

anxiety scores from the STAI range from 20 to 80, with higher scores indicating higher general proneness to anxiety. Normative data (Knight *et al.*, 1983; Spielberger *et al.*, 1983) suggest that most people from a healthy population score between 20 and 50 (mean score around 35); while scores above 50 may indicate some clinical relevance for an anxiety disorder (Fisher and Durham, 1999; Kennedy *et al.*, 2001; Julian, 2011).

If healthy volunteers who participate in fMRI studies exhibit lower anxiety levels than the general population, this could constrain the generalizability of fMRI data and have important implications for studies investigating processes associated with anxiety more specifically. For example, associations between brain responses and anxiety levels in healthy volunteers may not extend to the full range of anxiety scores typically observed in the general population. When applied to clinical studies, in-scanner effect sizes for differences between clinically anxious patients and controls may be overestimated, due to controls being abnormally “low” in anxiety compared to the average population estimate.

Initial support for our hypothesis of lower trait anxiety in fMRI study participants arose from preliminary pilot and published data from three studies (Caroline J. Charpentier *et al.*, 2016; C. J. Charpentier *et al.*, 2016; Charpentier *et al.*, 2018). Results from this preliminary dataset are summarized in **Table 1**. Trait anxiety was indeed lower in the fMRI study context than in the behaviour study context ( $T_{270}=2.679$ ,  $P=0.01$ , Cohen’s  $d=0.384$ ). There was no gender or age difference between study contexts, meaning those factors were unlikely to drive the observed difference in trait anxiety. However, the sample size ( $N=272$ ) was small (especially for the MRI context), and one factor that could be driving the difference in trait anxiety is whether participants were appropriately screened for psychiatric/affective disorders. In this preliminary sample, all fMRI subjects were screened, while a large proportion of the behaviour subjects ( $N=145$  out of 208) were not. In addition, all this data was collected by one experimenter at one institution, making it difficult to generalize.

Therefore, we set out to gather a large dataset of existing trait anxiety scores from behavioural and fMRI studies involving healthy volunteers across multiple institutions. In order to account for possible confounds and examine interaction effects, we additionally collected the following variables: gender, age, whether and how participants were screened for affective/psychiatric disorders, whether the study involved the presence of a stressor and/or pharmacological manipulation, whether the study was part of

anxiety research, the type of sample recruited, study duration, compensation rate, and whether the data was provided before or after participant exclusion.

[Insert Table 1]

Preliminary data	fMRI	Behaviour	Study Context Difference		
			Statistic	P-value	Effect size
N	64	208			
Gender: N <sub>F</sub> /N <sub>M</sub>	33/31	117/91	$\chi^2=0.435$	0.51	0.080
Trait anxiety ( $\pm$ SD)	34.422 ( $\pm 8.44$ )	38.226 ( $\pm 10.35$ )	$T_{270}=2.679$	0.01	0.384
Age ( $\pm$ SD)	25.891 ( $\pm 5.76$ )	24.995 ( $\pm 7.65$ )	$T_{270}=0.864$	0.39	0.124

**Table 1. Summary of preliminary dataset (N=272).** Independent, two-sample t-tests were run assuming unequal variance. Effect sizes for t-tests are Cohen's d values; and effect sizes for chi-square tests are standardized mean difference effect sizes calculated with the *esc\_chisq* function in R. For both types of effect sizes, 0.2 is considered a small effect, 0.5 a medium effect and 0.8 a large effect.

## Methods

**Procedure.** Trait anxiety total scores, from the State Trait Anxiety Inventory (Spielberger *et al.*, 1983) were obtained for a total of 3317 healthy adult participants (18 years and older) across 9 study sites and 5 countries: California Institute of Technology (USA), University of Maryland (USA), National Institute of Mental Health (USA), Universität Hamburg (Germany), Radboud University (the Netherlands), Leiden University (the Netherlands), University College London (UK), University of Oxford (UK), and University of Geneva (Switzerland). These excluded data from the preliminary dataset. A summary of the final dataset is provided in **Table 2**. Only data that previously collected in the different contributing labs was gathered; and data was completely de-identified before sharing. Possible duplicates – trait anxiety scores from the same participant in several different studies from the same lab – cannot be identified and are therefore not accounted for, although we expect the number of duplicates to be negligible. We asked labs to provide the following information along with trait anxiety scores: gender, age (in years), whether the study was a behavioural-only study or involved functional MRI scanning (study context), whether participants were appropriately screened for affective/psychiatric disorders (see **Table S2** for details of screening procedure), whether the study

involved the presence of a stressor and/or drug administration, and a short description of the study. The project was approved by the Caltech Institutional Review Board (minimal risk, exempt decision).

**Data analysis – mixed effect models.** Using the *lme4* package in R (Bates *et al.*, 2015), two mixed effects models were built (i) to examine the effect of study context (behaviour vs fMRI) while competing for variance with the other variables (Model 1) and (ii) to assess interaction between group and other variables (Model 2). Model 1 included fixed effects of study context, gender, age, psychiatric screening, stressor, and drug administration, as well as a fixed intercept and a random intercept (grouped by study site). Model 2 included the same effects as Model 1, with the addition of a random effect of study context (grouped by study site) and the following fixed interaction effects: context\*gender, context\*age, context\*psychiatric screening, context\*stressor and context\*drug administration. Study site was included as a random factor in all analyses given the variability in mean trait anxiety across study sites (**Table 3, All data** column). For both models, subjects with missing gender or age data (n=103) were excluded, and for Model 2, subjects from study sites that only provided data for one study context (n=173) were excluded to allow for the estimation of a random effect of condition for each study site. Model 1 thus included data from 3214 subjects, and Model 2 data from 3041 subjects. To determine the significance of individual effects, nested model comparison was performed, using Chi-square test in R (*anova* function) to compare the full model with the corresponding model lacking the one effect of interest. Effect sizes were obtained for pairwise comparisons of the marginal means using the *eff\_size* function from the *emmeans* package in R.

**Data analysis – Bayesian statistics.** Bayesian analyses were conducted using JASP (Love *et al.*, 2015) in order to provide support for the effects obtained with mixed effects models. Bayesian ANCOVA (Rouder and Morey, 2012; Rouder *et al.*, 2012) was used with trait anxiety as a dependent variable; study context, gender, psychiatric screening, stressor, and drug administration as fixed factors; age as a covariate; and study site as a random factor. To mirror the mixed effect analyses, two types of Bayesian model comparisons were performed. First, we compared pairs of models either including or not including a fixed effect of interest, with all other fixed effects included – this allowed determining the significance of main effects. Second, we compared pairs of models either including or not including an interaction effect of interest, with all fixed effects and all other interactions included. Note that only interactions with study context were considered. JASP's default prior was used. This pairwise model comparison allows drawing inference about which model best explains the data. In practice, the test generates a Bayes Factor (BF<sub>10</sub>), which represents the evidence for the full model relative to the null

model (which here simply lacks one effect of interest). The magnitude of  $BF_{10}$  was used to interpret the strength of evidence in favor of either model (Kass and Raftery, 1995; Jarosz and Wiley, 2014; Lee and Wagenmakers, 2014; Quintana and Williams, 2018). Evidence in favour of the model of interest was considered anecdotal ( $1 < BF_{10} < 3$ ), substantial ( $3 < BF_{10} < 10$ ), strong ( $10 < BF_{10} < 30$ ), very strong ( $30 < BF_{10} < 100$ ) or decisive ( $BF_{10} > 100$ ). Similarly, evidence in favour of the null model could also be qualified as anecdotal ( $0.33 < BF_{10} < 1$ ), substantial ( $0.1 < BF_{10} < 0.33$ ), strong ( $0.033 < BF_{10} < 0.1$ ), very strong ( $0.01 < BF_{10} < 0.033$ ) or decisive ( $BF_{10} < 0.01$ ).

**Follow-up analyses: (1) Effect of screening procedures.** To examine the role of specific psychiatric screening procedures in modulating trait anxiety differences between fMRI and behaviour study contexts, we repeated the analyses described above (mixed effect models and Bayesian tests), taking into account whether screening was performed by phone or in-person structured interview. The detailed screening procedures for each study site and study context are reported in **Table S2**. We also explored the distribution of trait anxiety scores for each type of screening procedure (no screening, phone screening, or full screening) and each study context, quantifying the mean and standard deviation (**Table 4**) as well as the mode and 80<sup>th</sup> percentile (**Fig. 4**) to characterize the distributions. **(2) Effect of state anxiety.** To assess whether the difference in trait anxiety observed between fMRI and behavioural studies could in fact be explained by a difference in state anxiety, we obtained state anxiety scores (STAI-S) for a subset of participants ( $N=2324$ ) from the main dataset. We ran a mixed level model on this subset of the data examining the effect of both study context and state anxiety (competing for variance) on trait anxiety. The model contained fixed effects of study context and state anxiety, as well as a random intercept for study site, and was compared with the same model excluding the fixed effect of study context.

**Data and code availability.** Data and code are available on the following github repository: [https://github.com/ccharpen/Trait\\_anxiety\\_MRI\\_BH](https://github.com/ccharpen/Trait_anxiety_MRI_BH), covered under a CC-BY-4.0 license.

[Insert Table 2]

Final data	fMRI	Behaviour	Study Context Difference		
			Statistic	P-value	Effect size
	N=1341	N=1976			
Gender: % female	51.2%	55.5%	$\chi^2=5.76$	0.016	0.083
Trait anxiety	35.772 ( $\pm 8.31$ )	37.820 ( $\pm 9.98$ )	$T_{3180}=6.41$	<0.0001	0.219

Age	24.135 ( $\pm 5.85$ )	25.638 ( $\pm 7.45$ )	$T_{3176.9}=6.40$	<0.0001	0.220
Screening: % yes	64.7%	63.2%	$\chi^2=0.852$	0.36	0.032
Stressor: % yes	50.3%	39.2%	$\chi^2=40.48$	<0.0001	0.222
Drug: % yes	13.9%	25.8%	$\chi^2=68.68$	<0.0001	0.291
Sample: % community	21.4%	27.2%	$\chi^2=480.9$	<0.0001	0.824
% convenience	68.8%	33.2%			
Anxiety research: % yes	13.4%	24.7%	$\chi^2=63.63$	<0.0001	0.280
All subjects included: % yes	88.2%	69.4%	$\chi^2=160.01$	<0.0001	0.450
Study duration in minutes	176.9 ( $\pm 131.5$ )	176.3 ( $\pm 174.6$ )	$T_{3279.4}=0.106$	0.92	0.004
Pay rate in USD per hour	25.97 ( $\pm 22.5$ )	17.25 ( $\pm 15.2$ )	$T_{2164.3}=12.42$	<0.0001	0.472

**Table 2. Summary of final dataset (N=3317).** For continuous variables, the table indicates mean values for each study context ( $\pm$  standard deviation) and results from independent, two-sample t-tests (assuming unequal variance). For discrete variables, percentages are shown. Trait anxiety, age and gender were obtained for each individual; the other variables display study-level characteristics. Effect sizes for t-tests are Cohen's d values; and effect sizes for chi-square tests are standardized mean difference effect sizes calculated with the *esc\_chisq* function in R. For both types of effect sizes, 0.2 is considered a small effect, 0.5 a medium effect and 0.8 a large effect.

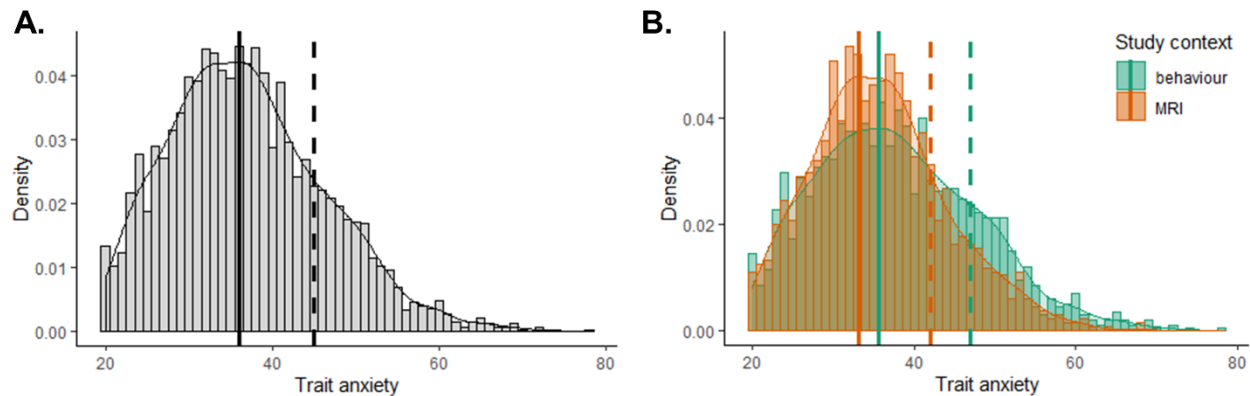
## Results

### Dataset summary and descriptive statistics

The distribution of trait anxiety scores is shown in **Figure 1**, across the entire sample (**Fig. 1A**) and separately for individuals participating in fMRI and behavioural studies (**Fig. 1B**). Note that the data only pertains to healthy volunteers and does not include any clinical samples. Mean trait anxiety across the entire sample was 36.99 ( $\pm 9.40$ ), consistent with normative data (Knight *et al.*, 1983; Spielberger *et al.*, 1983). Confirming our hypothesis and preliminary data, the difference in trait anxiety between fMRI and behavioural studies was also significant in the larger sample, albeit with a smaller, but non-negligible, effect size (t-test assuming unequal variance:  $T_{3180}=6.41$ ,  $P<0.0001$ ; Cohen's  $d=0.219$ ; **Table 2**). Interestingly, the distribution of trait anxiety scores across the two study contexts (**Fig. 1B**) indicates that the difference is driven by a larger proportion of individuals in fMRI studies scoring between 30 and 40, and a larger proportion of individuals in behavioural studies scoring above 45. While the difference in mean trait anxiety between study contexts was around 2 points on the trait anxiety scale, this difference rose to 5 points when examining the 80th percentile of the distribution. According to the standard scores provided in the scale manual (Spielberger *et al.*, 1983), this 5-point



difference suggests that the distribution of trait anxiety scores in fMRI studies is truncated by about 0.5 standard deviation compared to that in behavioural studies.



**Figure 1. Distribution of trait anxiety scores.** Density plots are shown, representing the proportion of the population at each trait anxiety score (bin = 1). Solid lines show the mode of the distribution; dashed lines the 80<sup>th</sup> percentile. (A) Distribution for the entire population (N=3317): mode=36.02, 80<sup>th</sup> percentile=45. (B) Separate distributions for behaviour (N=1976, green) and fMRI (N=1341, orange) study contexts, showing both lower mode (MRI=33.18, behaviour=35.68) and lower 80<sup>th</sup> percentile (MRI=42; behaviour=47) in the fMRI study context.

As observed in the preliminary data, it is possible that the difference in trait anxiety could be driven by one or several of the following factors, most of them found to be significantly different between study contexts (see **Table 2** for statistical inference). In the behaviour compared to fMRI context, participants were slightly older, pay rate was lower, and there was a higher proportion of female participants. More fMRI studies involved the presence of a stressor, recruited from convenience samples and provided trait anxiety data including all subjects (rather than only analysed subjects), whereas more behavioural studies involved drug administration and were part of anxiety research. However, the proportion of individuals that were clinically screened was not statistically different across study contexts, nor was the average study duration. Nonetheless, we performed follow-up analyses to regress out the variance explained by these possible confounds.

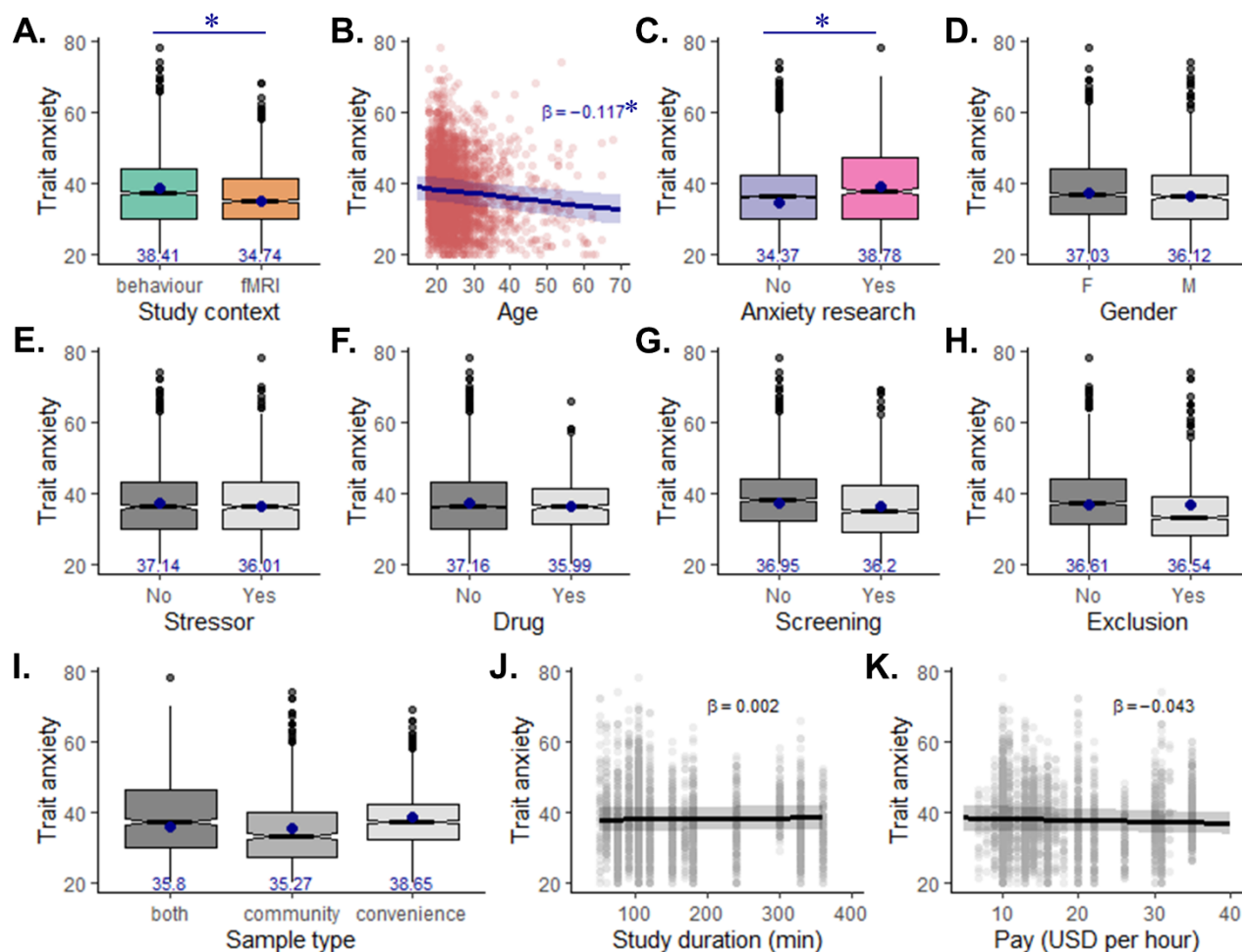
**The difference in trait anxiety between fMRI and behavioural studies is robust to potential confounds.**

Two analyses were performed to assess the effect of study context on trait anxiety while regressing out the variance explained by other possible confounding variables in the dataset: mixed effect modelling

and Bayesian ANCOVA (see **Methods** for details). Only results reaching threshold for both methods were considered robust enough to support our conclusions.

In a linear mixed effects model (Model 1), we included fixed effects of all factors (study context, screening, gender, age, stressor, drug administration, sample type, study duration, pay rate, anxiety research, and post-exclusion), as well as a fixed and random intercept for study site. We found a significant main effect of study context (estimate=-3.677 ±0.43 (SE),  $\chi^2=71.29$ ,  $P<0.0001$ ; **Fig. 2A**) with an effect size over the difference in marginal means of  $d=0.418$  (averaged over the levels of all other factors). This indicates that lower trait anxiety in individuals participating in fMRI over behavioural studies is a robust effect in our large sample, present even when competing for variance with multiple other factors such as gender, age, study details and recruitment strategy. In fact, accounting for the variance explained by these variables yielded a 90% larger effect size. Bayesian analyses supported this finding, with the model including all main effects outperforming the same model lacking only the effect of study context ( $BF_{10}>10^{14}$ ). This is indicative of decisive evidence for this effect. While the size of the effect is variable across the specific study sites that provided data for both contexts (medium to large effect in sites #1 and #9, small to medium effect in sites #3 and #7, negligible effect in sites #5, #6 and #8; **Table 3**), trait anxiety in all the sites was numerically lower for the fMRI context.

ACCEPTED MANUSCRIPT



**Figure 2. Main effects on trait anxiety (Model 1).** A mixed effects model was run to predict trait anxiety scores from eleven variables, all competing for variance: (A) study context (behaviour vs fMRI), (B) age, (C) whether the study was part of anxiety research, (D) gender, whether the study involved (E) a stressor, (F) a drug administration procedure, (G) psychiatric screening, (H) whether data was provided after participant exclusion, (I) sample type, (J) study duration in minutes, and (K) pay rate converted to USD per hour. Effects of categorical factors (A, C-I) are shown as box plots of the raw data; the blue dots and numbers represent the marginal means predicted from the model. Effects of continuous variables (B, J, K) are shown as scatter plots of trait anxiety as a function of the variable (dots: raw data; line: effect of the variable predicted by the model). The effects of study context, age, and anxiety research (A-C) were found to be significant both in the mixed effects model and using Bayesian tests (\*  $P < 0.001$  and  $BF_{10} > 100$ ).

[Insert Table 3]

Study site	All data		fMRI		Behaviour		Difference		
	N	Trait anxiety (± SD)	N	Trait anxiety (± SD)	N	Trait anxiety (± SD)	T	P	Effect size (d)
Site #1	255	38.35 (±11.19)	155	36.08 (±10.16)	100	41.85 (±11.84)	4.01	<0.001	0.53
Site #2	102	43.38 (±10.91)	0	-	102	43.38 (±10.91)	-	-	-
Site #3	890	36.12 (±7.90)	465	34.68 (±7.75)	425	37.69 (±7.78)	5.79	<0.001	0.39
Site #4	71	34.31 (±7.35)	0	-	71	34.31 (±7.35)	-	-	-
Site #5	100	35.95 (±8.13)	45	35.31 (±7.88)	55	36.47 (±8.37)	0.71	0.48	0.14
Site #6	440	39.30 (±6.95)	413	39.26 (±6.96)	27	39.81 (±6.93)	0.40	0.69	0.08
Site #7	94	29.66 (±5.71)	61	29.02 (±5.32)	33	30.85 (±6.29)	1.42	0.16	0.32
Site #8	441	34.28 (±9.40)	55	33.84 (±9.56)	386	34.34 (±9.39)	0.37	0.72	0.053
Site #9	924	38.02 (±10.53)	147	32.78 (±8.12)	777	39.01 (±10.65)	8.09	<0.001	0.61

**Table 3. Data summary by study site.** Sample sizes and mean trait anxiety scores (± standard deviation) are reported for each site, for all data and separately for the fMRI and Behaviour study contexts. Statistics for the difference between fMRI and Behaviour contexts are also reported in the right-most column, specifically T and p-value from two-tailed independent sample t-tests (unequal variance) and effect size using Cohen's d.

### Lower trait anxiety with age and in studies focusing on anxiety research

We then set out to analyse the effect of other variables on trait anxiety to determine which effects are robust to the other variables in the model. The mixed effects model (Model 1) revealed significant effect of age (lower trait anxiety in older individuals: estimate=-0.117 ±0.025 (SE),  $\chi^2=21.56$ ,  $P<0.0001$ ; **Fig. 2B**) and of anxiety research (higher trait anxiety in studies that are part of anxiety research: estimate=4.416 ±0.71 (SE),  $\chi^2=38.54$ ,  $P<0.001$ ; **Fig. 2C**). Both were supported by the Bayesian test with decisive evidence (age:  $BF_{10}=4528$ ; anxiety research:  $BF_{10}>10^7$ ). We note that the negative correlation between age and trait anxiety is robust to outliers (excluding individuals over age 50:  $R_{3161}=-0.136$ ; excluding individuals over age 35:  $R_{3001}=-0.141$ ; both  $P<0.001$ ).

Evidence for effects of gender, stressor and drug administration was mixed, as the mixed effects model suggested significant fixed effects (higher trait anxiety in females: estimate=-0.907 ±0.32 (SE),

$\chi^2=7.773$ ,  $P=0.0053$ , **Fig. 2D**; lower trait anxiety in studies involving a stressor: estimate= $-1.137 \pm 0.49$  (SE),  $\chi^2=5.283$ ,  $P=0.022$ , **Fig. 2E**; lower trait anxiety in studies involving drug administration: estimate= $-1.171 \pm 0.57$  (SE),  $\chi^2=4.212$ ,  $P=0.04$ , **Fig. 2F**). However, the Bayesian analyses only indicated anecdotal evidence (gender:  $BF_{10}=1.83$ ; stressor:  $BF_{10}=0.645$ , drug administration:  $BF_{10}=0.392$ ).

Finally, both analyses showed no significant effect of psychiatric screening (estimate= $-0.747 \pm 0.51$  (SE),  $\chi^2=2.082$ ,  $P=0.149$ ; **Fig. 2G**), exclusion (estimate= $-0.063 \pm 0.66$  (SE),  $\chi^2=0.009$ ,  $P=0.92$ , **Fig. 2H**), sample type (community vs both estimate= $-0.535 \pm 0.69$  (SE), convenience vs both estimate= $2.843 \pm 1.73$  (SE),  $\chi^2=3.327$ ,  $P=0.190$ ; **Fig. 2I**), or study duration (estimate= $0.002 \pm 0.001$  (SE),  $\chi^2=1.785$ ,  $P=0.182$ , **Fig. 2J**), pay rate (estimate= $-0.043 \pm 0.033$  (SE),  $\chi^2=1.435$ ,  $P=0.231$ , **Fig. 2K**) on trait anxiety. Bayesian tests suggested strong to anecdotal evidence for these null effects (exclusion:  $BF_{10}=0.075$ , sample type:  $BF_{10}=0.18$ , screening:  $BF_{10}=0.21$ , study duration:  $BF_{10}=0.30$ , pay rate:  $BF_{10}=0.51$ ).

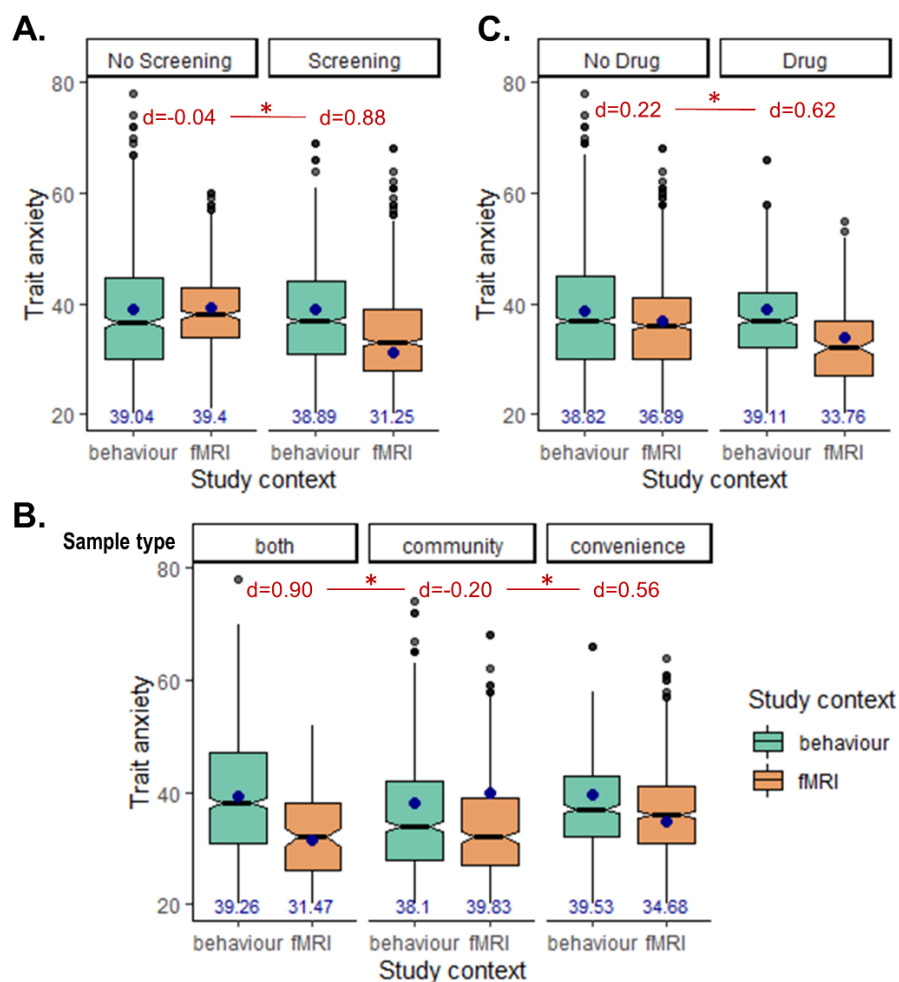
### **Behaviour-fMRI trait anxiety differences are modulated by screening, drug administration, and sample type**

We then examined whether the difference in trait anxiety between behavioural and fMRI studies was moderated by any of the other variables. To test this, we built a second mixed-effects model (Model 2) which, in addition to Model 1 effects, included the following 2-way interactions with study context as fixed effects: context\*gender, context\*age, context\*screening, context\*stressor, context\*drug, context\*sample type, context\*study duration, and context\*pay rate. A random effect of study context (with study site as random variable) was also included, allowing to model the interaction between context and site. Note that (i) this model only included the seven study sites that had data from both behaviour and fMRI study contexts, thus leading to a slightly reduced sample size of 3041, and (ii) given the small proportion of studies (especially fMRI studies) that were classified as anxiety research and provided data after participant exclusions, we were unable to include the interaction between study context and these variables. Finally, because the difference between behavioural and fMRI studies was our main question of interest, we did not investigate interactions between the other factors (i.e. not including study context).

We found significant interactions between study context and screening (estimate= $-8.008 \pm 2.71$  (SE),  $\chi^2=7.860$ ,  $P=0.005$ ; **Fig. 3A**), between study context and sample type (estimate= $9.525 \pm 3.81$  (SE),

$\chi^2=6.730$ ,  $P=0.035$ ; **Fig. 3B**) and between study context and drug administration (estimate=-3.414  $\pm$ 1.46 (SE),  $\chi^2=4.933$ ,  $P=0.026$ ; **Fig. 3C**). All three interactions were supported by the Bayesian tests (context\*screening:  $BF_{10}=13.07$ ; context\*sample type:  $BF_{10}=3.282$ ; context\*drug:  $BF_{10}=4.742$ ). The context\*screening interaction was such that higher trait anxiety in behaviour compared to fMRI study contexts was only present when subjects were screened (effect size of difference in marginal means  $d=0.88$ ) compared to when they weren't screened ( $d=-0.042$ ). The context\*sample type interaction was such that the fMRI-behaviour difference in anxiety was present in studies using convenience samples ( $d=0.558$ ) or a mix of convenience and community samples ( $d=0.898$ ) but not in studies relying on community samples only ( $d=-0.199$ ). Finally, the context\*drug interaction was such that the fMRI-behaviour difference in trait anxiety was larger in studies involving a drug administration procedure ( $d=0.62$ ) than in those without ( $d=0.22$ ). All other interactions were not significant ( $\chi^2<1.4$ ,  $P>0.24$ ), as supported by the Bayesian tests ( $BF_{10}<0.45$ ). Given that some categories had no data (e.g. combined stress and drug administration study without screening), we refrained from investigating higher-level interactions than the ones reported above.

ACCEPTED MANUSCRIPT



**Figure 3. Moderating factors of the behaviour-fMRI difference in trait anxiety (Model 2).** Interaction effects with study context were added to the mixed effects model and the interactions with (A) screening, (B) sample type, and (C) drug administration, were found to be significant. Effects are shown as box plots of the raw data; the blue dots and numbers represent the marginal means from the interaction effect predicted by the model; the numbers in red represent the effect sizes associated with the behaviour-fMRI differences in marginal means.

### Post-hoc analysis: effect of screening type

In the analyses reported above, participants were considered screened for affective/psychiatric disorders if either a phone screening or in-person structured interview was conducted; and not screened if absence of psychiatric condition was based purely on self-report of meeting eligibility criteria specified in the recruitment material or if no such eligibility criteria were specified. However, it is likely that the exact type of screening procedure (see **Table S2** for details) may differ across study contexts and play more of a modulatory role on trait anxiety scores. To examine this, we ran follow-up analyses in which instead of a binary variable, screening was classified into one of three types: no screening, phone screening, or full in-person screening. Numbers and mean trait anxiety for each

screening type and study context are reported in **Table 4**, including the breakdown for those specific sites that used the same screening procedure across both study contexts. We found that the proportions of participants screened by phone, in person, or not screened did not differ across study contexts ( $\chi^2=2.21$ ,  $P=0.33$ ).

Re-running linear mixed effect model 1, but distinguishing between phone and full screening procedures, showed that the difference in trait anxiety across study contexts remained significant (estimate= $-3.388 \pm 0.43$  (SE),  $\chi^2=60.21$ ,  $P<0.0001$ ,  $BF_{10}>10^{11}$ ). There was also a significant main effect of psychiatric screening type ( $\chi^2=41.24$ ,  $P<0.0001$ ,  $BF_{10}>10^6$ , **Fig 4A**), with higher trait anxiety for unscreened compared to fully screened individuals (estimate= $2.653 \pm 0.60$  (SE)), and for individuals screened by phone compared to those that screened in person (estimate= $5.168 \pm 0.82$  (SE)). Re-running linear mixed model 2, testing for interactions with study context, revealed a significant interaction between study context and the type of screening procedure ( $\chi^2=23.54$ ,  $P<0.0001$ ,  $BF_{10}=335.5$ ). Mean trait anxiety scores collapsed across all sites (**Table 4**) showed that the interaction was driven by lower trait anxiety for fMRI relative to behaviour contexts when phone screening procedures were used ( $T_{1245}=10.46$ ,  $P<0.001$ ,  $d=0.57$ ) but not for studies with no screening ( $T_{1198.7}=-0.58$ ,  $P=0.56$ ,  $d=0.031$ ) or studies with full in-person screening ( $T_{688.74}=0.64$ ,  $P=0.52$ ,  $d=0.047$ ).

Finally, examining the distribution of trait anxiety scores across study contexts and screening procedures (**Fig. 4B**) revealed some interesting findings. First, while there was no difference in mean trait anxiety between behaviour and fMRI study contexts for unscreened individuals (**Fig. 4B top**), the distributions exhibit several differences: the mode is lower for behavioural studies (33.43 vs 37.64), while the 80<sup>th</sup> percentile is lower for fMRI studies (45 vs 48), confirming the narrower distribution of trait anxiety scores in fMRI studies when no psychiatric screening is performed at recruitment. For individuals screened by phone (**Fig. 4B middle**), both the mode (32.28 vs 36.15) and 80<sup>th</sup> percentile (40.2 vs 48) were lower in fMRI study contexts, driven by a smaller proportion of individuals scoring above 42. When individuals were fully screened using an in-person structured clinical interview (**Fig. 4B bottom**), the two distributions matched almost exactly between study contexts (mode: behaviour=29.72, fMRI=30.40; 80<sup>th</sup> percentile: behaviour=41, fMRI=40).

We also note that specific sites could be driving some of the differences between no screening and phone screening (**Table 4** and **Fig. 4C**). Specifically, Site #9 played an important contribution to the difference observed in the case of phone screening (with Site #3 exhibiting a smaller but significant

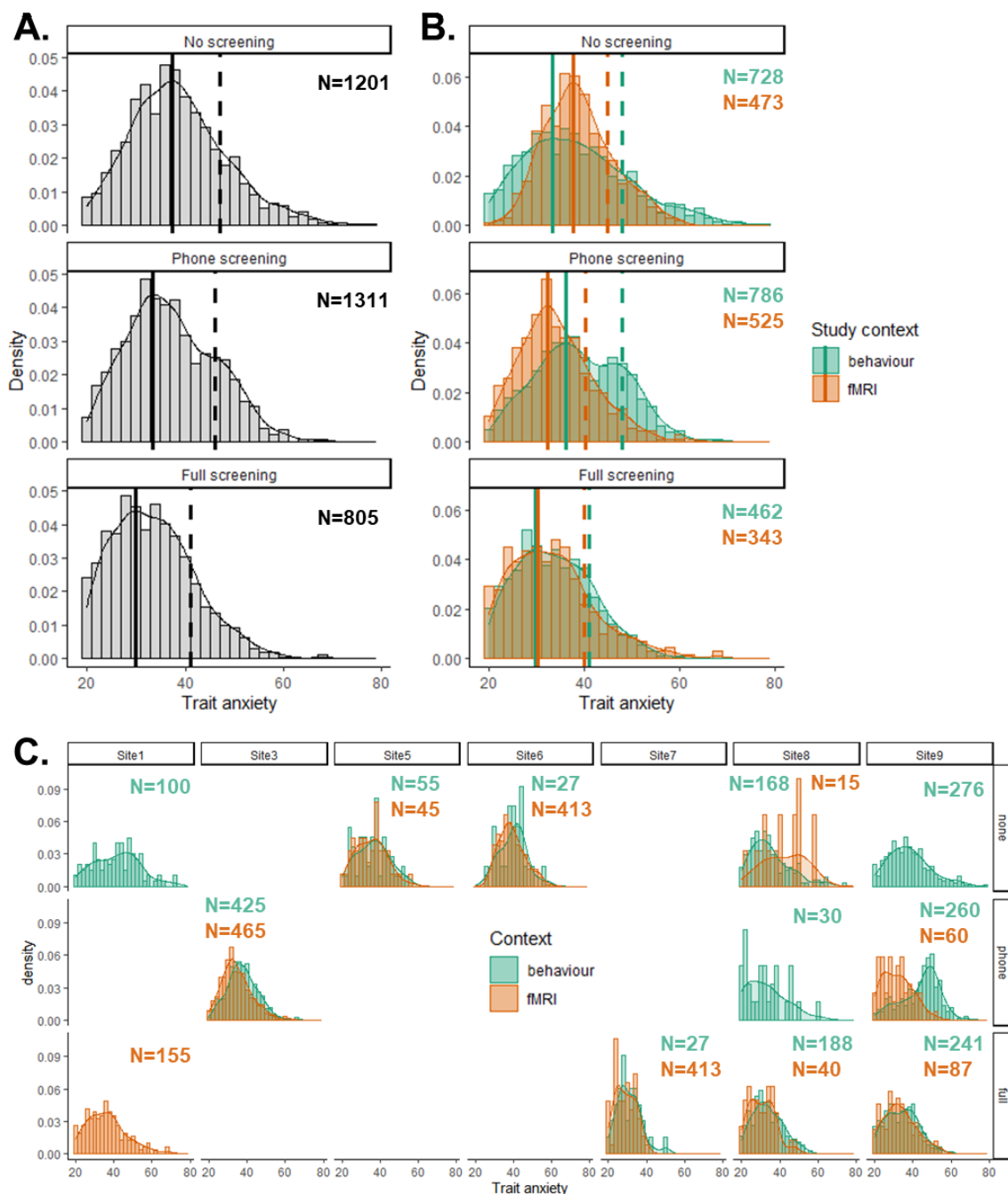


effect in the same direction). In the absence of screening, however, we see that Site #8 actually shows an effect in the opposite direction, with the caveat that the sample size for the fMRI group in Site #8 is extremely small (N=15), making the comparison for this particular site very underpowered and difficult to interpret. Overall, this heterogeneity between sites seems reduced in the case of full screening, for which trait anxiety scores are consistent across study contexts in all three sites that provided data for this arm (i.e. Sites #7, #8 and #9).

[Insert Table 4]

Screening type	Site	Behaviour		fMRI		Difference		
		N	Trait anxiety (± SD)	N	Trait anxiety (± SD)	T	P	d
No screening	<i>all</i>	<b>728</b>	<b>38.68 (±11.08)</b>	<b>473</b>	<b>38.99 (±7.30)</b>	<b>-0.58</b>	<b>0.56</b>	<b>0.031</b>
	#5	55	36.47 (±8.37)	45	35.31 (±7.88)	0.71	0.48	0.14
	#6	27	39.81 (±6.93)	413	39.26 (±6.96)	0.40	0.69	0.079
	#8	168	35.21 (±10.62)	15	42.40 (±10.99)	-2.51	0.01	0.68
Phone	<i>all</i>	<b>786</b>	<b>39.33 (±9.31)</b>	<b>525</b>	<b>34.36 (±7.78)</b>	<b>10.46</b>	<b>&lt;0.001</b>	<b>0.57</b>
	#3	425	37.69 (±7.78)	465	34.68 (±7.75)	5.78	<0.01	0.39
	#9	260	44.13 (±9.90)	60	31.92 (±7.72)	8.94	<0.01	1.28
Full	<i>all</i>	<b>462</b>	<b>33.90 (±8.09)</b>	<b>343</b>	<b>33.50 (±9.04)</b>	<b>0.64</b>	<b>0.52</b>	<b>0.047</b>
	#7	33	30.85 (±6.29)	61	29.02 (±5.32)	1.49	0.14	0.33
	#8	188	33.80 (±7.93)	40	30.63 (±6.67)	2.36	0.02	0.41
	#9	241	34.39 (±8.36)	87	33.37 (±8.38)	0.97	0.33	0.12

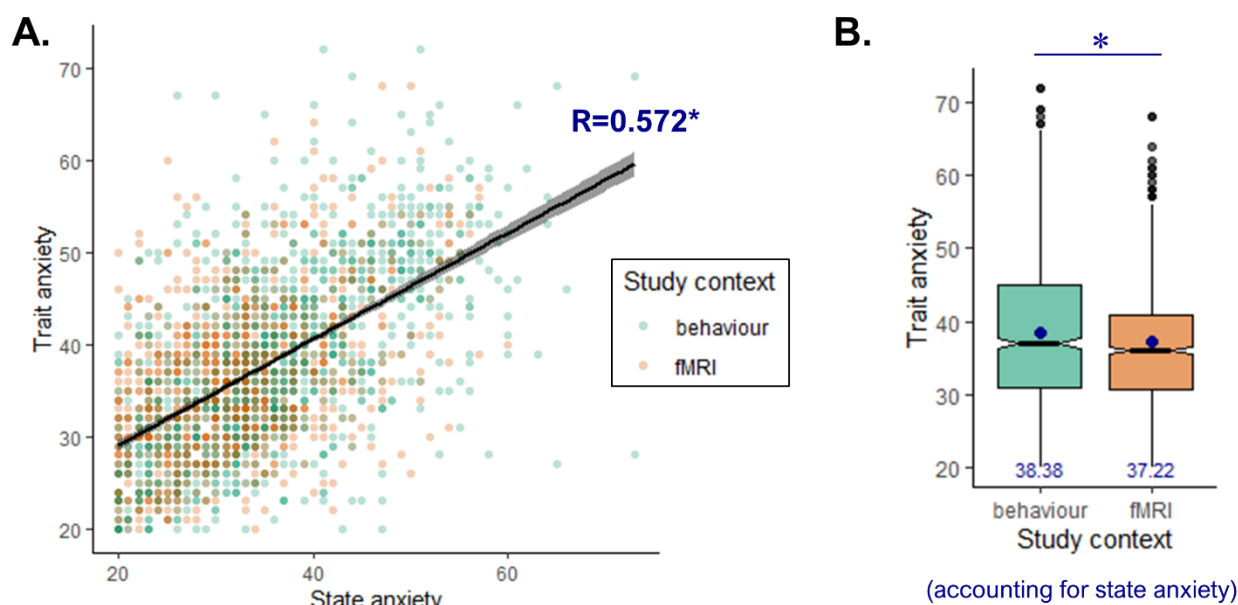
**Table 4. Trait anxiety across study contexts and screening procedures.** The number of individuals, as well as mean trait anxiety and standard deviation, are shown separately for each screening procedure (no screening, phone screening, full in-person screening) and each study context (behaviour, fMRI). Numbers in bold and italics are for the entire dataset, collapsing across all study sites. The breakdown for the specific sites in which the same procedure was used for both study contexts is also shown.



**Figure 4. Distribution of trait anxiety scores split by screening procedure.** Density plots of trait anxiety scores are shown (bin=2), separately for individuals who were not screened for psychiatric/affective disorders (top panels), screened by phone (middle panels) or fully screened with an in-person structured clinical interview (bottom panels). Numbers of participants included in each distribution are shown above each density plot. (A) Distribution for the entire population. (B) Separate distributions for behaviour and fMRI study contexts. Solid lines in A and B show the mode of the distribution; dashed lines the 80<sup>th</sup> percentile. (C) Separate distributions across study contexts and across sites. Only sites that provided trait anxiety scores for at least one behavioural study and one fMRI study are included.

### Relationship with state anxiety

Given that trait and state anxiety scores are generally highly correlated, it is possible that the observed difference in trait anxiety between behavioural and fMRI studies is in fact driven by state anxiety scores. To assess this possibility, we gathered state anxiety scores for a subset of the entire dataset (N=2324) across 5 sites with both behavioural and fMRI studies (**Fig. S1**). The correlation between state and trait anxiety in these individuals was indeed high ( $R=0.572$ ,  $P<0.001$ , **Fig. 5A**) and there was a significant difference in state anxiety between study contexts ( $T_{2289,6}=9.59$ ,  $P<0.001$ , **Fig S1B**). Nonetheless, the amount of unshared variance between the two variables (67.3%) was sufficient to examine the effect of study context on trait anxiety while regressing out the variance explained by state anxiety scores. To do so, we ran a final mixed effects model, which predicted trait anxiety from study context and state anxiety (see **Methods** for details). We found that the effect of study context on trait anxiety remained significant (estimate=-1.162  $\pm$ 0.39 (SE),  $\chi^2=8.848$ ,  $P=0.003$ ,  $BF_{10}=6.04$ ; **Fig. 5B**), and thus could not be explained by state anxiety alone.



**Figure 5. Dissociating state and trait anxiety.** (A) We obtained state anxiety scores in a subset of the data (N=2324) and plot trait anxiety as a function of state anxiety for each of these individuals, as well as the best-fitting regression line. (B) The difference in trait anxiety between behavioural and fMRI studies remains significant when regressing out the variance explained by state anxiety.

## Discussion

In this study, we provide substantial evidence, in a large-scale dataset of healthy participants across multiple sites, that individuals participating in fMRI studies exhibit on average lower levels of trait anxiety than individuals participating in behavioural studies only. We show that this effect is even stronger when regressing out the variance in trait anxiety explained by multiple other factors, such as age, gender, but also recruitment strategies (sample type, psychiatric screening) and other study details (presence of a stressor or drug, study duration, pay rate, anxiety research). In addition, the trait anxiety difference remained significant when state anxiety scores were included in the model. Both mixed modelling approaches and Bayesian analyses supported this effect. Interestingly, while the mean difference in trait anxiety scores was relatively small (2 point-difference, effect size=0.219), we note that the effect size nearly doubled (4-point difference in the marginal means,  $d=0.418$ , **Fig. 2A**) when variance due to other factors was regressed out, indicative of a moderate effect size. Additionally and importantly, distributions across study contexts markedly differed from each other, with a much narrower and somewhat truncated distribution of trait anxiety scores in fMRI studies relative to behavioural studies. A recent study reported similar distributions of trait anxiety scores when comparing their behavioural and fMRI samples (Sjouwerman *et al.*, 2020). This difference in the distributions has two main consequences. First, it suggests that non-clinical fMRI samples are less representative of the general population than non-clinical behavioural samples. Second, the narrower range will make examining individual differences in trait anxiety more difficult in fMRI compared to behavioural studies.

Procedures in place to screen participants for psychiatric and/or affective disorders were found to modulate the distribution of trait anxiety scores in different ways for fMRI and behavioural study contexts. Specifically, when no or minimal (i.e. phone) screening was performed, the range of trait anxiety scores was narrower in fMRI compared to behavioural studies, while the two distributions matched when full in-person clinical interviews were used. Several possible factors could explain the observed differences. For studies employing phone or online screening procedures, participants with higher anxiety might be screened out of fMRI studies more often than behavioural studies, because of inherent differences in screening questions. For example, during recruitment of fMRI studies, participants are likely asked additional screening questions, such as history of claustrophobia, which would usually not be asked for behavioural studies. It is also possible that participants are more likely to not reveal, or not be aware of, their history of psychiatric disorders when participating in a behavioural study for which screening does not occur in person. Finally, a self-selection bias during recruitment is also possible, whereby individuals with high trait anxiety are less likely to volunteer to

participate in fMRI studies, even if they meet all eligibility criteria. When full in-person clinical screening is performed, however, we believe that participants with higher anxiety are more likely to be excluded from the study, given the high comorbidity between many DSM-5 disorders (for which meeting criteria will usually result in exclusion from a healthy control sample) and elevated anxiety. This is irrespective of whether the study involves neuroimaging or not. Undergoing MRI scanning has been found to be anxiogenic, because of claustrophobia, discomfort, and/or fear of learning about potential incidental findings (Meléndez and McCrank, 1993; Katz *et al.*, 1994; Murphy and Brunberg, 1997); therefore, anxious individuals are likely to find the experience of MRI scanning more aversive and elect not to participate. While excluding participants with claustrophobia from fMRI studies may partly explain the bias (Katz *et al.*, 1994; Murphy and Brunberg, 1997), whether other specific components of anxiety play a role remains unclear. Factor analyses of the STAI (Bieling *et al.*, 1998; Vigneau and Cormier, 2008; Wang *et al.*, 2018) suggested different components of trait anxiety, such as anxiety-present vs anxiety-absent components (corresponding to items reflecting negative vs positive emotional experiences) or components assessing anxiety, worry, sadness, self-deprecation, as well as general negative affect. Whether a subset of these components is more likely to weigh in on the decision to take part in an fMRI or behavioural study remains an open question for future investigation. We note this analysis was beyond the scope of the present study, given that individual item scores from the trait anxiety questionnaires were not obtained in the data.

Examining the distributions of trait anxiety scores across sites, screening procedures and study context (**Fig. 4C**) indicates substantial between-sites heterogeneity in how screening procedures may modulate the behaviour-fMRI trait anxiety difference. This raises the possibility that there is still a lot of unexplained variance between sites. Such heterogeneity is likely due to the observational nature of the study - analysing existing data rather than carefully controlling variables between sites to allow for robust comparisons and quantification of interaction effects. Therefore, site differences should be interpreted with caution since any inferred cause for these differences is likely to be speculative. Instead, we hope that future studies will rely on carefully controlled designs or experimental manipulations to empirically address whether and how participants' decisions to sign up for a study and researchers' decisions to include participants are influenced by the screening procedure, specific questions asked during screening, the recruitment materials, or the participant's level of anxiety during sign-up. This would allow determining whether the sampling bias arises before or after screening. Nevertheless, the present findings are important and robust to those site-specific effects since our

regression model accounts for variance between sites, suggesting that trait anxiety is lower in fMRI compared to behavioural studies over and beyond the differences observed between sites.

Our results also revealed that the type of sample or participant pool subjects are recruited from seems to matter, consistent with previous evidence suggesting an effect of sample composition on neuroimaging findings (LeWinn *et al.*, 2017). Specifically, the difference in trait anxiety between fMRI and behavioural studies was larger in studies that relied at least partly on convenience samples than in studies recruiting from the community. Finally, the effect was also larger in studies involving a drug administration procedure (i.e., pharmacology studies), suggesting that the sampling or self-selection bias towards individuals with low trait anxiety is more evident in studies combining fMRI with pharmacology (compared to fMRI only). Furthermore, our findings speak to other factors that explain some of the variance in individual trait anxiety scores. We found a negative correlation between age and trait anxiety, consistent with past literature suggesting trait anxiety decreases with age (Knight *et al.*, 1983; Nakazato and Shimonaka, 1989; Regier *et al.*, 1990). The evidence for an effect of gender on trait anxiety, however, was mixed. Consistent with the literature suggesting both higher prevalence of anxiety disorders (McLean *et al.*, 2011) and higher self-reported anxiety (Knight *et al.*, 1983; Spitzer *et al.*, 2006) in females than males, we also report higher trait anxiety in females. This effect was significant in the mixed effects model but was not robustly supported by Bayesian tests. We also found higher trait anxiety in studies that were considered part of anxiety research, possibly because these studies might mention their relevance to anxiety research in recruitment materials and therefore be more likely to appeal to participants experiencing more anxiety.

While the large scale of the present dataset allowed us to ensure the robustness of the effects, with data obtained from multiple institutions and regressing out the effects of multiple potential confounds, we note possible limitations. First, contributing institutions were mostly located in the USA and northern Europe, thus leaving open the possibility that the observed effects may not generalize to data collected in other parts of the world. Second, the variables we included in the analyses (age, gender, screening type, sample type, stressor, drug, study duration, pay rate, study site, anxiety research and exclusion) are of course not exhaustive and one could imagine that other mediators are likely to explain additional variance in trait anxiety scores and/or in the willingness to participate in fMRI studies (Leach *et al.*, 2008). Examples include socio-economic status, race/ethnicity, urban living, ruminative and depressive states, neuroticism, physical health, remuneration, or other components of the study design. Collecting these additional variables would not have been possible in the current dataset given that they were either not measured in the first place or would have compromised the anonymization of the dataset.

Finally, we found the difference in trait anxiety scores was found to be robust to state anxiety in a subset of the data; however, we do not discard a possible role of state anxiety in the self-selection bias as well. Similarly, recent literature suggests that trait anxiety may not exclusively measure anxiety *per se*, but rather reflect negative affect more generally (Hur *et al.*, 2019; Knowles and Olatunji, 2020), both in its behavioural and neurobiological signatures (Shackman *et al.*, 2016). Whether the difference observed between fMRI and behavioural participants is specific to trait anxiety or relates to general negative affect thus remains an open question.

Overall, the finding of lower trait anxiety, as well as narrower distribution of trait anxiety scores, in fMRI compared to behavioural studies has implications for both previously published and future research in the field of cognitive neuroscience as a whole, and for anxiety research more specifically. These differences may be responsible for failed replications, whereby a behavioural effect of interest, and/or a moderating effect of trait (or induced/state) anxiety, evidenced in a behavioural study fails to replicate in a follow-up fMRI study (e.g (Bolton and Robinson, 2017; Garibbo *et al.*, 2019)) or vice versa. Because of the narrower range of trait anxiety values in fMRI studies, this may also enhance the differences between patient and control groups in studies of psychiatric populations, whereby control subjects have lower trait anxiety than the general population. While the present findings may carry some relevance for clinical studies, we note that the dataset did not include any trait anxiety scores from clinically anxious individuals; therefore, we do not know whether the observed difference between behavioural and fMRI study contexts would expand to clinical samples. It is possible that the bias may not actually be present in clinical studies, given that patients' motivations or benefits for participating in research may be different and lead to greater recruitment into fMRI studies than in non-clinical samples. This intriguing possibility warrants further investigation.

Taken together, these findings point towards possible recommendations for cognitive neuroscience researchers who run both fMRI and behavioural studies to measure individual differences in anxiety and carefully consider and mitigate potential sources of recruitment bias. Our finding that distributions of trait anxiety scores between fMRI and behavioural studies match almost perfectly when full in-person psychiatric screening interviews are conducted suggests that such screening procedure may be one way to ensure similar levels of trait anxiety across study contexts. This is particularly relevant for researchers recruiting from convenience samples (i.e. undergraduate students) or running pharmacology studies, both of which were associated with larger trait anxiety differences between behavioural and fMRI contexts. However, while screening can ensure better matching of trait anxiety across study contexts, it may still lead to samples that are not representative of the general population,

as anxious individuals may be excluded from all study contexts at a higher rate. This could result in a loss of power to examine individual differences and undermine relevance for psychopathology.

A solution would then be to rely on methods that help recruit participants with higher anxiety into fMRI studies, such as the use of a mock scanner, virtual reality, or psychological interventions. These methods have been successful in alleviating MRI anxiety in pediatric populations (Viggiano *et al.*, 2015) and patients with anxiety (Garcia-Palacios *et al.*, 2007; Tugwell *et al.*, 2018), and could therefore be expanded to the general population to reduce sampling biases. Using stratified sampling, whereby trait anxiety or dispositional negativity is measured at screening in a large sample after which study participants are selectively recruited from that sample to ensure representation across the full range, would also help mitigating the observed bias (Hur *et al.*, 2020).

In conclusion, our recommendations in light of the present findings are as follows. First, regardless of the specific causes behind this bias, this study sheds light on the possibility that fMRI samples are less representative of the general population than behavioural samples, or at the very least that behavioural and fMRI samples are different from each other, when it comes to trait anxiety. This is likely to be problematic for mechanistic or experimental research, irrespective of generalizability. Second, researchers should adopt recruitment and/or screening strategies that can help them mitigate this bias if it is likely to impact the validity or interpretation of their results. Finally, future research should further explore possible causes of this bias and mitigation strategies, particularly through controlled experiments. Probing more deeply into individual reasons for participating in fMRI studies and differences in screening procedures seems necessary to ensure researchers can enforce a distribution of psychological and psychiatric profiles that is representative of the general population.

## **Funding:**



CJC is supported by a Wellcome Trust Sir Henry Wellcome Postdoctoral Fellowship (218642/Z/19/Z). JOD and LKP are supported by the Caltech Conte Center for the Neurobiology of Social Decision-Making (NIMH award: P50MH094258). JOD is supported by a National Institute on Drug Abuse R01 DA040011. OJR and YY were funded by A Medical Research Foundation Equipment Competition Grant (C0497, Principal Investigator OJR), and a Medical Research Council Career Development Award to OJR (MR/K024280/1). AM is supported by Wellcome Trust-NIH PhD Studentship. LS was supported by the German Research Foundation (DFG; grants: SFB/TRR 58, project B10; SCHW1357/14-1; SCHW1357/16-1). EP and AG are supported by the Swiss Centre for Affective Sciences. JPR and VV are supported by the Wellcome Trust (101798/Z/13/Z). VV is supported by a NIHR UCLH Biomedical Research Centre Fellowship. QJMH acknowledges support by the Max Planck Society and the UCL NIHR BRC. MB is supported by a MRC Clinician Scientist Fellowship (MR/N008103/1) and by the NIHR Oxford Health Biomedical Research Centre. CJH and SEM are supported by the NIHR Oxford Health Biomedical Research Centre. LP was supported by the National Institute of Mental Health (R01 MH071589 and R01 MH112517). The funders had no role in study design, data analysis, decision to publish or preparation of the manuscript.

### Author contributions:

CJC, OJR, HWC, JPR and JOD were involved in the conception of the study. CJC, PF, ERP, VL, MST, LMK, AF, YY, NL, AM, VV, QJMH, KAM, VK, FK, AC, GZ, FMK, NW, AG, EP, SM, AK, MB, LKP collected and contributed data to the large-scale dataset. CJC performed data analysis and wrote the manuscript. PF, ERP, VL, MST, LMK, AF, NL, VV, QJMH, KAM, LKP, RC, KR, LP, CJH, HWC, CG, LS, OJR, JOD commented on the manuscript and contributed to the interpretation of the results and to the revisions.

### Competing interests:

OJR's current MRC senior fellowship is partially in collaboration with Cambridge Cognition (who plan to provide in kind contribution). He is running an investigator initiated trial with medication donated by Lundbeck (escitalopram and placebo, no financial contribution), holds an MRC-Proximity to discovery award with Roche (in kind contributions, sponsor travel), and has completed consultancy work for Peak and IESO digital health. JPR has performed consultancy work for Cambridge Cognition, Takeda and GE Healthcare. MB has received travel expenses from Lundbeck for attending conferences, and has acted as a consultant for J&J and for CHDR. CJH has received consultancy fees from P1vital Ltd, Sage Pharmaceuticals, Zogenix, J&J and Pfizer. These disclosures are made in the interest of full transparency and do not constitute a direct conflict of interest with the current work.

### References

Bar-Haim, Y., Lamy, D., Pergamin, L. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin*, **133**, 1–24

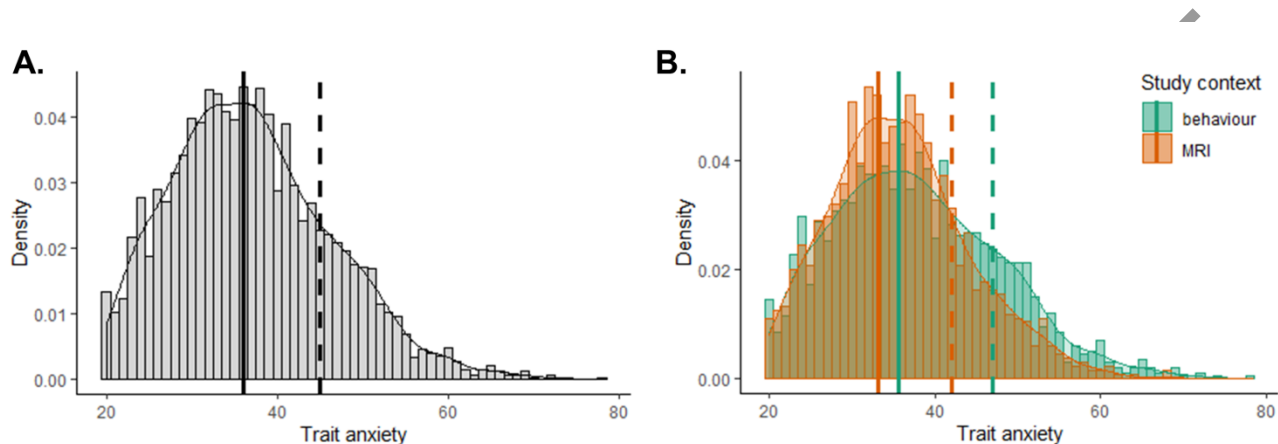
- Bates, D., Maechler, M., Bolker, B., et al. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1–48
- Beddington, J., Cooper, C.L., Goswami, U., et al. (2008). The mental wealth of nations. *Nature*, **455**, 1057–59
- Bieling, P.J., Antony, M.M., Swinson, R.P. (1998). The state-trait anxiety inventory, trait version: Structure and content re-examined. *Behaviour Research and Therapy*
- Bishop, S.J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience*, **12**, 92–98
- Bolton, S., Robinson, O.J. (2017). The impact of threat of shock-induced anxiety on memory encoding and retrieval. *Learning & Memory*, **24**, 532–42
- Browning, M., Behrens, T.E., Joham, G., et al. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci*, **18**, 590–96
- Carlisi, C.O., Robinson, O.J. (2018). The role of prefrontal–subcortical circuitry in negative bias in anxiety: Translational, developmental and treatment perspectives. *Brain and Neuroscience Advances*, **2**, 239821281877422
- Charpentier, C.J., Aylward, J., Roiser, J.P., et al. (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biological Psychiatry*, **81**, 1014–22
- Charpentier, C.J., Bromberg-Martin, E.S., Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, **115**, E7255–64
- Charpentier, Caroline J., De Martino, B., Sim, A.L., et al. (2016). Emotion-induced loss aversion and striatal-amygdala coupling in low-anxious individuals. *Social Cognitive and Affective Neuroscience*, **11**, 569–79
- Charpentier, C. J., De Neve, J.-E., Li, X., et al. (2016). Models of affective decision making: how do feelings predict choice? *Psychological Science*, **27**, 763–75
- Cisler, J.M., Koster, E.H.W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, **30**, 203–16
- Clark, L., Li, R., Wright, C.M., et al. (2012). Risk-avoidant decision making increased by threat of electric shock. *Psychophysiology*, **49**, 1436–43
- Engel-Yeger, B., Dunn, W. (2011). The relationship between sensory processing difficulties and anxiety level of healthy adults. *British Journal of Occupational Therapy*, **74**, 210–16
- Etkin, A., Klemenhagen, K.C., Dudman, J.T., et al. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron*, **44**, 1043–55
- Etkin, A., Prater, K.E., Hoefft, F., et al. (2010). Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *Am J Psychiatry*, **167**, 545–54
- Farmer, A.S., Kashdan, T.B. (2012). Social anxiety and emotion regulation in daily life: spillover effects on positive and negative social events. *Cogn Behav Ther*, **41**, 152–62
- Fisher, P.L., Durham, R.C. (1999). Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the STAI-T across outcome studies since 1990. *Psychological Medicine*, **29**, 1425–34
- Garcia-Palacios, A., Hoffman, H., Richards, T., et al. (2007). Use of Virtual Reality Distraction to Reduce Claustrophobia Symptoms during a Mock Magnetic Resonance Imaging Brain Scan: A Case Report. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, **10**, 485–88
- Garibbo, M., Aylward, J., Robinson, O.J. (2019). The impact of threat of shock-induced anxiety on the neural substrates of memory encoding and retrieval. *Social cognitive and affective neuroscience*, **14**, 1087–96
- Gossett, E.W., Wheelock, M.D., Goodman, A.M., et al. (2018). Anticipatory stress associated with functional magnetic resonance imaging: Implications for psychosocial stress research. *International Journal of Psychophysiology*, **125**, 35–41
- Grillon, C. (2002). Startle reactivity and anxiety disorders: aversive conditioning, context, and neurobiology. *Biological Psychiatry*, **52**, 958–75
- Grillon, C., Robinson, O.J., Cornwell, B., et al. (2019). Modeling anxiety in healthy humans: a key intermediate bridge between basic and clinical sciences. *Neuropsychopharmacology*, **0**, 1–12
- Hommel, B., Fischer, R., Colzato, L.S., et al. (2012). The effect of fMRI (noise) on cognitive control. *Journal of*

- Experimental Psychology: Human Perception and Performance*, **38**, 290–301
- Hur, J., DeYoung, K.A., Islam, S., et al. (2020). Social context and the real-world consequences of social anxiety. *Psychological Medicine*, **50**, 1989–2000
- Hur, J., Stockbridge, M.D., Fox, A.S., et al. (2019). Chapter 16 - Dispositional negativity, cognition, and anxiety disorders: An integrative translational neuroscience framework. In: N. B. T.-P. in B. R. Srinivasan (ed). *Emotion and Cognition*. Elsevier, p. 375–436.
- Jarosz, A.F., Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *Journal of Problem Solving*, **7**, 2–9
- Julian, L.J. (2011). Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care and Research*, **63**, 467–72
- Kass, R.E., Raftery, A.E. (1995). Bayes Factors. *J Am Stat Assoc*, **90**, 773–95
- Katz, R.C., Wilson, L., Frazer, N. (1994). Anxiety and its determinants in patients undergoing Magnetic Resonance Imaging. *Journal of Behavior Therapy and Experimental Psychiatry*, **25**, 131–34
- Kennedy, B.L., Schwab, J.J., Morris, R.L., et al. (2001). Assessment of State and Trait Anxiety in Subjects with Anxiety and Depressive Disorders. *Psychiatric Quarterly*, **72**, 263–76
- Knight, R.G., Waal-Manning, H.J., Spears, G.F. (1983). Some norms and reliability data for the State-Trait Anxiety Inventory and the Zung Self-Rating Depression scale. *British Journal of Clinical Psychology*, **22**, 245–49
- Knowles, K.A., Olatunji, B.O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the State-Trait Anxiety Inventory. *Clinical Psychology Review*, **82**, 101928
- Kobald, S.O., Getzmann, S., Beste, C., et al. (2016). The impact of simulated MRI scanner background noise on visual attention processes as measured by the EEG. *Scientific Reports*, **6**, 1–10
- Leach, L.S., Christensen, H., Mackinnon, A.J., et al. (2008). Gender differences in depression and anxiety across the adult lifespan: the role of psychosocial mediators. *Social Psychiatry and Psychiatric Epidemiology*, **43**, 983–98
- Lee, M., Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: a practical course*. Cambridge: Cambridge University Press.
- LeWinn, K.Z., Sheridan, M.A., Keyes, K.M., et al. (2017). Sample composition alters associations between age and brain structure. *Nature Communications*, **8**, 874
- Love, J., Selker, R., Marsman, M., et al. (2015). JASP (Version 0.7.1) [Computer Software]
- Maner, J.K., Richey, J.A., Cromer, K., et al. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences*, **42**, 665–75
- McLean, C.P., Asnaani, A., Litz, B.T., et al. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *Journal of psychiatric research*, **45**, 1027–35
- Meléndez, J.C., McCrank, E. (1993). Anxiety-Related Reactions Associated With Magnetic Resonance Imaging Examinations. *JAMA*, **270**, 745–47
- Mkrтчian, A., Aylward, J., Dayan, P., et al. (2017). Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. *Biological Psychiatry*, **82**, 532–39
- Murphy, K.J., Brunberg, J.A. (1997). Adult claustrophobia, anxiety and sedation in MRI. *Magnetic Resonance Imaging*, **15**, 51–54
- Nakazato, K., Shimonaka, Y. (1989). The Japanese State-Trait Anxiety Inventory: Age and Sex Differences. *Perceptual and Motor Skills*, **69**, 611–17
- Poli, E., Angrilli, A. (2015). Greater general startle reflex is associated with greater anxiety levels: A correlational study on 111 young women. *Frontiers in Behavioral Neuroscience*, **9**, 1–6
- Quintana, D.S., Williams, D.R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry*, **18**, 178
- Regier, D.A., Narrow, W.E., Rae, D.S. (1990). The epidemiology of anxiety disorders: The epidemiologic catchment area (ECA) experience. *Journal of Psychiatric Research*, **24**, 3–14
- Robinson, O.J., Krinsky, M., Lieberman, L., et al. (2014). The dorsal medial prefrontal (anterior cingulate) cortex-amygdala aversive amplification circuit in unmedicated generalised and social anxiety disorders: an observational study. *Lancet Psychiatry*, **1**, 294–302
- Robinson, O.J., Pike, A.C., Cornwell, B., et al. (2019). The translational neural circuitry of anxiety. *Journal of*

- Neurology, Neurosurgery and Psychiatry*, **90**, 1353–60
- Robinson, O.J., Vytal, K., Cornwell, B.R., et al. (2013). The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Front Hum Neurosci*, **7**, 203
- Rouder, J., Morey, R. (2012). Default Bayes factors for model selection in regression. *Multivar Behav Res*, **47**, 877–903
- Rouder, J.N., Morey, R.D., Speckman, P.L., et al. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, **56**, 356–74
- Shackman, A.J., Sarinopoulos, I., Maxwell, J.S., et al. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion*, **6**, 40–61
- Shackman, A.J., Tromp, D.P.M., Stockbridge, M.D., et al. (2016). Dispositional negativity: An integrative psychological and neurobiological perspective. *Psychological bulletin*, **142**, 1275–1314
- Sjouwerman, R., Scharfenort, R., Lonsdorf, T.B. (2020). Individual differences in fear acquisition: multivariate analyses of different emotional negativity scales, physiological responding, subjective measures, and neural activation. *Scientific Reports*, **10**, 15283
- Skouras, S., Gray, M., Critchley, H., et al. (2013). fMRI scanner noise interaction with affective neural processes. *PLoS ONE*, **8**, 1–13
- Spielberger, C.D., Gorsuch, R.L., Lushene, P.R., et al. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., et al. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, **166**, 1092–97
- Stein, M.B., Simmons, A.N., Feinstein, J.S., et al. (2007). Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. *Am J Psychiatry*, **164**, 318–27
- Tugwell, J.R., Goulden, N., Mullins, P. (2018). Alleviating anxiety in patients prior to MRI: A pilot single-centre single-blinded randomised controlled trial to compare video demonstration or telephone conversation with a radiographer versus routine intervention. *Radiography*, **24**, 122–29
- Viggiano, M.P., Giganti, F., Rossi, A., et al. (2015). Impact of Psychological Interventions on Reducing Anxiety, Fear and the Need for Sedation in Children Undergoing Magnetic Resonance Imaging. *Pediatric reports*, **7**, 5682
- Vigneau, F., Cormier, S. (2008). The Factor Structure of the State-Trait Anxiety Inventory: An Alternative View. *Journal of Personality Assessment*, **90**, 280–85
- Wang, T., Li, M., Xu, S., et al. (2018). The Factorial Structure of Trait Anxiety and Its Mediating Effect Between Mindfulness and Depression. *Frontiers in psychiatry*, **9**, 514
- Yao, N., Chen, S., Qian, M. (2018). Trait anxiety is associated with a decreased visual working memory capacity for faces. *Psychiatry Research*, **270**, 474–82

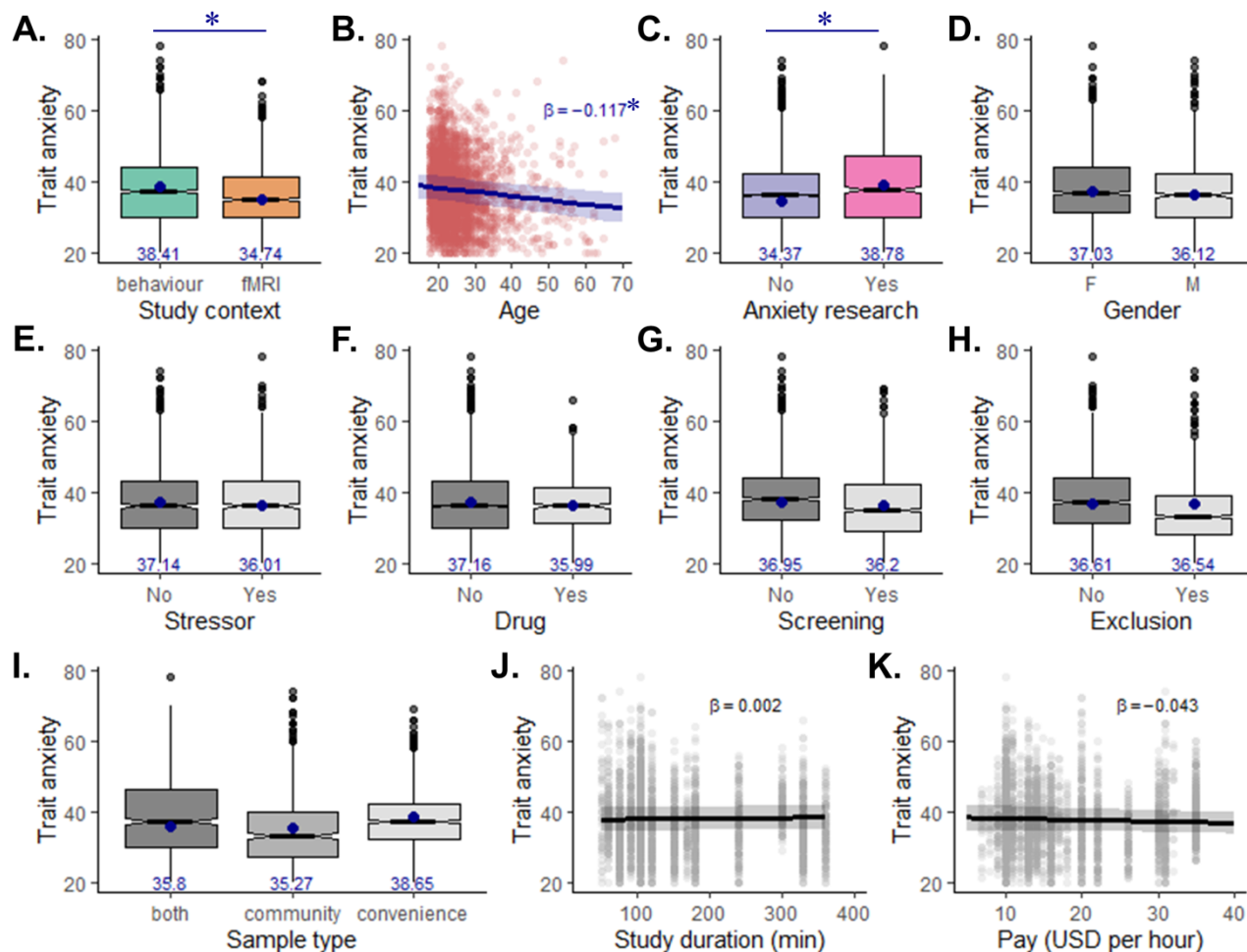
## Figure Legends:

**Figure 1. Distribution of trait anxiety scores.** Density plots are shown, representing the proportion of the population at each trait anxiety score (bin = 1). Solid lines show the mode of the distribution; dashed lines the 80<sup>th</sup> percentile. (A) Distribution for the entire population (N=3317): mode=36.02, 80<sup>th</sup> percentile=45. (B) Separate distributions for behaviour (N=1976, green) and fMRI (N=1341, orange) study contexts, showing both lower mode (MRI=33.18, behaviour=35.68) and lower 80<sup>th</sup> percentile (MRI=42; behaviour=47) in the fMRI study context.

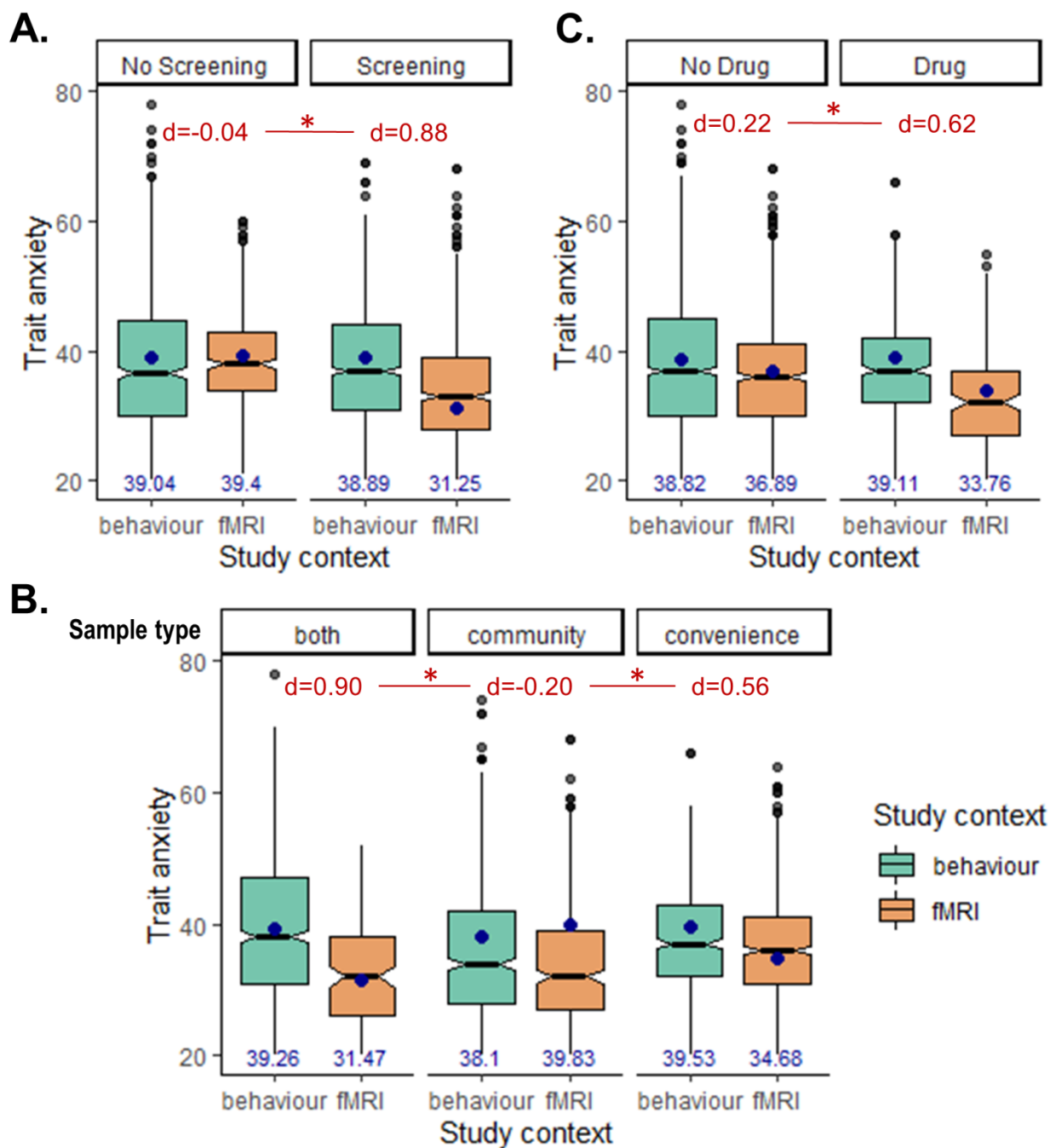


**Figure 2. Main effects on trait anxiety (Model 1).** A mixed effects model was run to predict trait anxiety scores from eleven variables, all competing for variance: (A) study context (behaviour vs fMRI), (B) age, (C) whether the study was part of anxiety research, (D) gender, whether the study involved (E) a stressor, (F) a drug administration procedure, (G) psychiatric screening, (H) whether data was provided after participant exclusion, (I) sample type, (J) study duration in minutes, and (K) pay rate converted to USD per hour. Effects of categorical factors (A, C-I) are shown as box plots of the raw data; the blue dots and numbers represent the marginal means predicted from the model. Effects of continuous variables (B, J, K) are shown as scatter plots of trait anxiety as a function of the variable (dots: raw data; line: effect of the variable predicted by the model). The effects of study context, age, and anxiety research (A-C) were found to be significant both in the mixed effects model and using Bayesian tests (\*  $P < 0.001$  and  $BF_{10} > 100$ ).

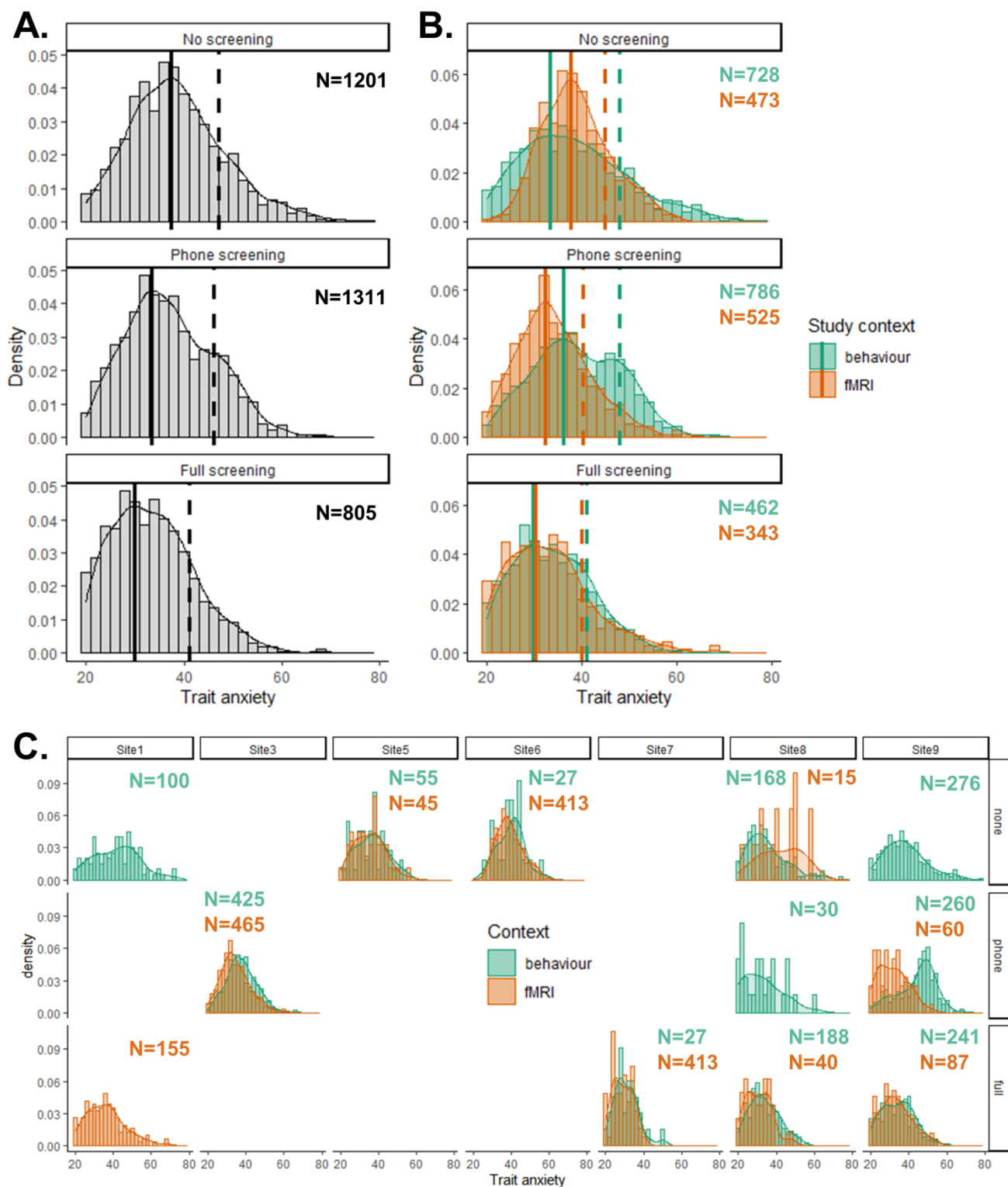
ACCEPTED MANUSCRIPT



**Figure 3. Moderating factors of the behaviour-fMRI difference in trait anxiety (Model 2).** Interaction effects with study context were added to the mixed effects model and the interactions with (A) screening, (B) sample type, and (C) drug administration, were found to be significant. Effects are shown as box plots of the raw data; the blue dots and numbers represent the marginal means from the interaction effect predicted by the model; the numbers in red represent the effect sizes associated with the behaviour-fMRI differences in marginal means.

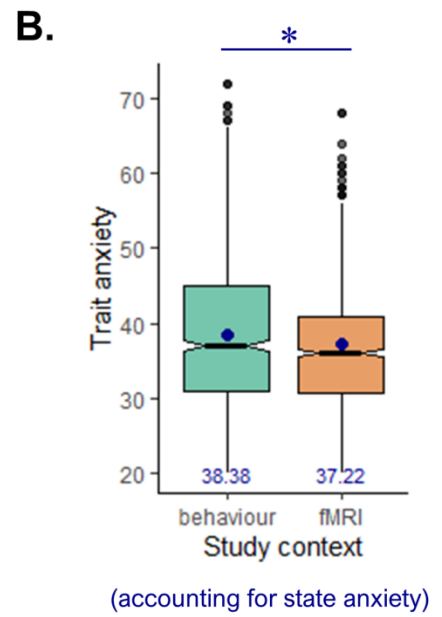
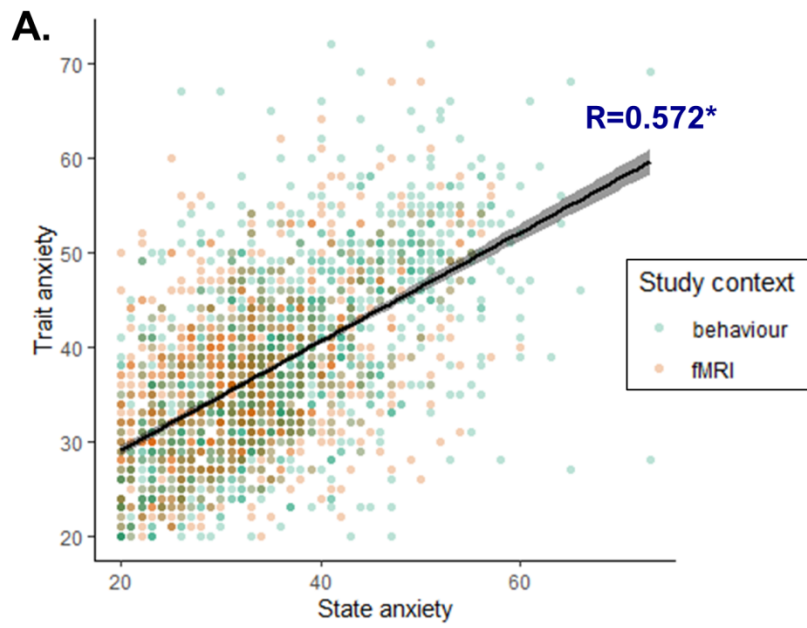


**Figure 4. Distribution of trait anxiety scores split by screening procedure.** Density plots of trait anxiety scores are shown (bin=2), separately for individuals who were not screened for psychiatric/affective disorders (top panels), screened by phone (middle panels) or fully screened with an in-person structured clinical interview (bottom panels). Numbers of participants included in each distribution are shown above each density plot. (A) Distribution for the entire population. (B) Separate distributions for behaviour and fMRI study contexts. Solid lines in A and B show the mode of the distribution; dashed lines the 80<sup>th</sup> percentile. (C) Separate distributions across study contexts and across sites. Only sites that provided trait anxiety scores for at least one behavioural study and one fMRI study are included.



**Figure 5. Dissociating state and trait anxiety.** (A) We obtained state anxiety scores in a subset of the data (N=2324) and plot trait anxiety as a function of state anxiety for each of these individuals, as well as the best-fitting regression line. (B) The difference in trait anxiety between behavioural and fMRI studies remains significant when regressing out the variance explained by state anxiety.





ACCEPTED MANUSCRIPT