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

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

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

*Corresponding author contact information: ccharpen@caltech.edu

Abstract

Over the past three decades, MRI has become a key tool to study how cognitive processes are implemented in the human brain. However, the question of whether participants recruited into MRI 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, if a behavioural effect discovered in a non-imaging context does not replicate in a neuroimaging environment. Here, we tested the hypothesis, motivated by preliminary findings (n=272), that MRI study participants differ from behaviour-only study participants on one fundamental individual difference variable: trait anxiety. Analysing a large-scale dataset drawn from multiple institutions (n=3317) and controlling for possible confounding variables, we found robust support for lower trait anxiety in MRI study participants, consistent with a sampling bias. Distributions of trait anxiety scores differed most markedly when psychiatric screening was minimal. Our findings highlight the need for surveying trait anxiety at recruitment and for appropriate screening procedures, in an attempt to mitigate this 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 MRI 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 MRI 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[1] fail to replicate[2] inside the scanning environment. Similarly, some effects may be more easily found in MRI 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[3–5], can affect cognitive and affective processes, their neural basis, and hormonal responses[6], 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 particular 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 MRI studies exhibit lower trait anxiety than individuals who participate in behavioural studies. Even 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[7–9]. While perhaps not unexpected, the hypothesis of lower trait anxiety in MRI 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[10], anxiety is also a normative adaptive function that varies across the 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[11,12]. Myriad studies have suggested that a wide range of cognitive functions are modulated by anxiety levels (see [13] for a review): sensory processing and gating[14–16], attentional biases toward negative emotional stimuli[17,18], decreased emotion regulation[19,20], deficits in attentional control[21], reduced working memory performance[22,23], impairments during reinforcement learning[24,25] and increased risk avoidance during decision-making[26–28]. Neuroimaging studies have provided evidence for heightened amygdala responses to negative emotional stimuli[29,30] and reduced connectivity between the prefrontal cortex and the amygdala[19,31,32] 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[33] (STAI). Trait anxiety scores from the STAI range from 20 to 80, with higher scores indicating higher general proneness to anxiety. Normative data[33,34] 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[35–37].

If individuals who participate in MRI studies exhibit lower anxiety levels than the general population, this could impose constraints on 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 MRI study participants arose from a preliminary dataset comprising pilot and published data from three studies[38–40]. Results from this preliminary dataset are summarized in **Table 1**. Trait anxiety was indeed lower in the MRI 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 MRI subjects were screened, while a large proportion of the behaviour subjects (N=145 out of 208) were

not. In addition, all of 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 labs across multiple institutions who routinely collect trait anxiety measures in their behavioural and MRI studies[1,25,41–54]. In order to control 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, and whether the study involved the presence of a stressor and/or pharmacological manipulation (see **Methods** for details).

Preliminary data	**MRI**	**Behaviour**	**Study Context Difference**		
			Statistic	**P-value**	**Effect size**
N	64	208			
Gender: N_F/N_M	33/31	117/91	χ^2=0.435	0.51	0.080
Trait anxiety (± SD)	34.422 (±8.44)	38.226 (±10.35)	T_{270}=2.679	0.01	0.384
Age (± SD)	25.891 (±5.76)	24.995 (±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.

Results

Dataset summary and descriptive statistics

The final dataset included data from 3317 healthy volunteers across 9 different study sites across Europe and the USA, excluding data from the preliminary dataset. A summary of the final dataset is provided in **Table 2**. The distribution of trait anxiety scores is shown in **Figure 1**, across the entire sample (**Fig. 1A**) and separately for individuals participating in MRI and behavioural studies (**Fig. 1B**). Mean trait anxiety across the entire sample was 36.99 (±9.40), consistent with normative data[33,34]. Confirming our hypothesis and preliminary data, the difference in trait anxiety between MRI 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 MRI 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.

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, all found to be significantly different between study contexts (see **Table 2** for statistical inference). In the behaviour compared to MRI context, participants were slightly older (Behaviour: 25.64 years; MRI: 24.14 years), there was a higher proportion of female participants (Behaviour: 55.5% females; MRI: 51.2% females), a higher proportion of participants in studies involving drug administration (Behaviour: 25.8%; MRI: 13.9%), and a lower proportion of participants in studies involving a stressor (Behaviour: 39.2%; MRI: 50.3%). However, the proportion of individuals that were clinically screened was not statistically different across study contexts (Behaviour: 63.2%; MRI: 64.7%). Nonetheless, we performed follow-up analyses to control for these possible confounds.

Final data	**MRI**	**Behaviour**	**Study Context Difference**		
			Statistic	**P-value**	**Effect size**
N	1341	1976			
Gender: N_F/N_M	687/654	1096/880	$\chi^2=5.76$	0.016	0.083
Trait anxiety (± SD)	35.772 (±8.31)	37.820 (±9.98)	$T_{3180}=6.41$	<0.0001	0.219
Age (± SD)	24.135 (±5.85)	25.638 (±7.45)	$T_{3176.9}=6.40$	<0.0001	0.220
Screening: N_{YES}/N_{NO}	868/473	1248/728	$\chi^2=0.852$	0.36	0.032
Stressor: N_{YES}/N_{NO}	675/666	774/1202	$\chi^2=40.48$	<0.0001	0.222
Drug: N_{YES}/N_{NO}	186/1155	510/1466	$\chi^2=68.68$	<0.0001	0.291

Table 2. Summary of final dataset (N=3317). 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.

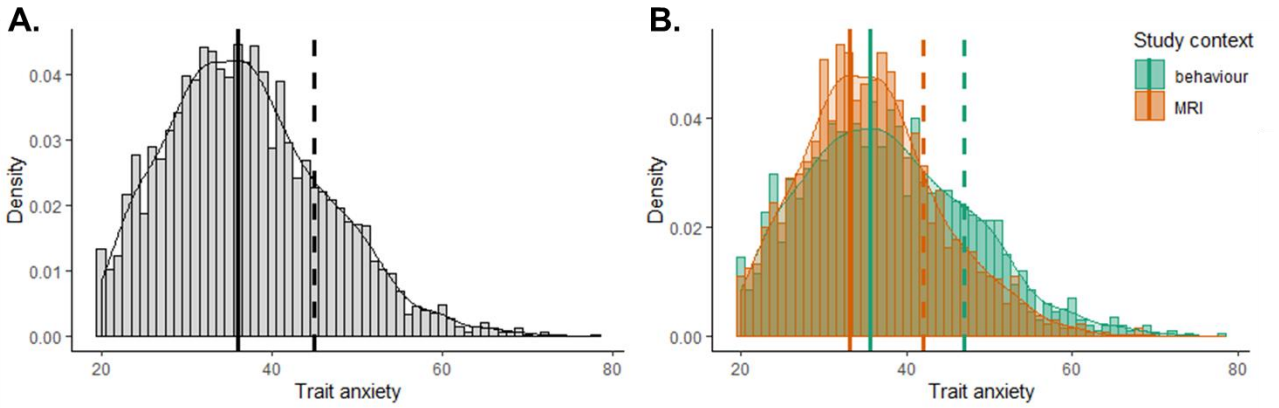

Figure 1. Distribution of trait anxiety scores. Density plots are shown, representing the proportion of the population at each trait anxiety score (bin = 1). Solid lines show the mode of the distribution; dashed lines the 80th percentile. (**A**) Distribution for the entire population (N=3317): mode=36.02,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

Trait anxiety difference between MRI and behavioural studies is robust to potential confounds.

Two analyses were performed to assess the effect of study context on trait anxiety while controlling for other variables in the dataset: mixed effect modelling using *lme4* package in R[55], as well as Bayesian ANCOVA in JASP[56,57]. Only results reaching threshold for both methods were considered robust enough to support our conclusions. Given the variability in mean trait anxiety across study sites (**Table 3**, **All data** column), we included study site as a random factor in all analyses.

In a linear mixed effects model (Model 1, see **Methods** for details), we included fixed effects of all factors (study context, screening, gender, age, stressor, drug administration), as well as a fixed and random intercept. We found a significant main effect of study context (estimate=-3.609 ±0.42 (SE), χ^2=73.44, P<0.0001; **Fig. 2A**) with an effect size over the difference in marginal means of d=0.406 (averaged over the levels of all other factors). This indicates that lower trait anxiety in individuals participating in MRI over behavioural studies is a robust effect in our large sample, present even when controlling for gender, screening, age, and the presence of a stressor or drug administration. In fact, accounting for the variance explained by these variables yielded an 85% larger effect size. Bayesian analyses confirmed 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 conditions (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 MRI condition.

	All data		MRI		Behaviour		Difference		
Study site	**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 MRI and Behaviour study contexts. Statistics for the difference between MRI 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 screening and age

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 a significant effect of age (lower trait anxiety in older individuals: estimate=-0.140 ±0.025 (SE), χ^2=32.33, P<0.0001; **Fig. 2B**), confirmed with decisive evidence by the Bayesian test ($BF_{10}>10^5$). Evidence for an effect of gender was mixed, as the mixed effects model suggested a significant fixed effect (higher trait anxiety in females: estimate=-0.940 ±0.33 (SE), χ^2=8.320, P=0.0039; **Fig. 2C**) but the Bayesian analyses only indicated anecdotal evidence (BF_{10}=2.788). Finally, both analyses showed no significant effect of psychiatric screening (estimate=-0.668 ±0.51 (SE), χ^2=1.681, P=0.195; **Fig. 2D**), stressor (estimate=0.486 ±0.41 (SE), χ^2=1.418, P=0.234; **Fig. 2E**) or drug administration (estimate=-0.729 ±0.49 (SE), χ^2=2.220, P=0.136; **Fig. 2F**) on trait anxiety, with the Bayesian test suggesting substantial

evidence for null effects (screening: BF_{10}=0.129, stressor: BF_{10}=0.117, drug administration: BF_{10}=0.205).

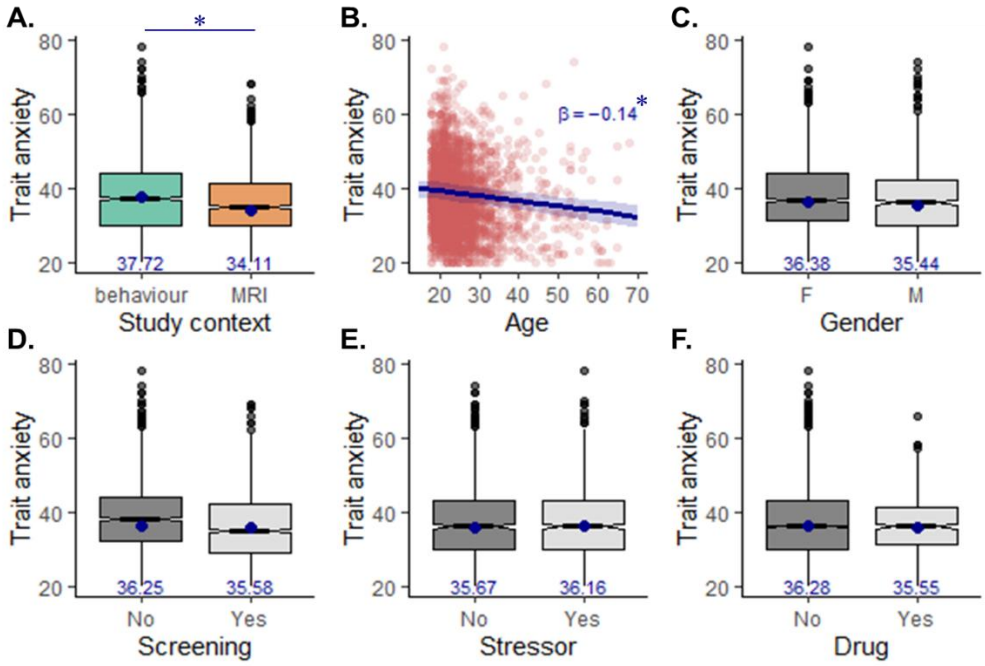

Figure 2. Main effects on trait anxiety (Model 1). A mixed effects model was run to predict trait anxiety scores from six variables: (**A**) study context (behaviour vs MRI), (**B**) age, (**C**) gender, (**D**) psychiatric screening, (**E**) stressor, and (**F**) drug administration. Effects of categorical factors (**A, C-F**) are shown as box plots of the raw data; the blue dots and numbers represent the marginal means predicted from the model. (**B**) Scatter plot of trait anxiety as a function of age (red dots: raw data; blue line: effect of age predicted by the model). The effects of study context and age (**A, B**) were found to be significant both in the mixed effects model and using Bayesian tests (* P<0.001 and BF_{10}>100), whereas the effects of screening, stressor, drug and gender (**C-F**) were not.

Behaviour-MRI trait anxiety differences are modulated by study context and age

Our final analysis examined whether the difference in trait anxiety between behavioural and MRI studies was moderated by any of the control variables. To test this, we built a second mixed-effects model (Model 2) which, in addition to Model 1 effects, included all 2-way interactions with study context (context*gender, context*age, context*screening, context*stressor, context*drug) as fixed effects, as well as a random effect of study context (with site as random variable). Note that this model only included the seven study sites that had data from both behaviour and MRI study contexts, thus leading to a slightly reduced sample size of 3041. Because the difference between behavioural and MRI 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=-4.719 ±1.29 (SE), χ^2=12.46, P=0.0004; **Fig. 3A**), between study context and age (estimate=0.164 ±0.055 (SE), χ^2=8.80, P=0.003; **Fig. 3B**) and between study context and stressor (estimate=-2.592 ±0.80 (SE), χ^2=9.844, P=0.0017; **Fig. 3C**). All three interactions were corroborated with Bayesian tests (context*screening: BF_{10}=46.73; context*age: BF_{10}=11.93; context*stressor: BF_{10}=3.86). The context*screening interaction was such that higher trait anxiety in behaviour compared to MRI study contexts was only present when subjects were screened (effect size of difference in marginal means d=0.54) compared to when they weren't screened (d=0.007). The context*age interaction revealed a negative correlation between age and trait anxiety only in behavioural studies (β=-0.168) but not in MRI studies (β=-0.003). The context*stressor interaction was such that the MRI-behaviour difference in trait anxiety was larger in studies involving a stressor (d=0.42) than in studies without a stressor (d=0.13). Finally, there was no evidence for a context*drug interaction (estimate=-1.016 ±1.13 (SE), χ^2=0.740, P=0.39; **Fig. 3D**) or context*gender interaction (estimate=-0.373 ±0.66 (SE), χ^2=0.319, P=0.57; **Fig. 3E**), consistent with the Bayesian tests (context*drug: BF_{10}=0.24; context*gender: BF_{10}=0.034). For completeness, mean trait anxiety, standard deviation and sample size are reported in **Table S1**, broken down by each of the five categorical factors (study context, gender, psychiatric screening, stressor, and drug administration). 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.

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 (χ^2=2.21, P=0.33).

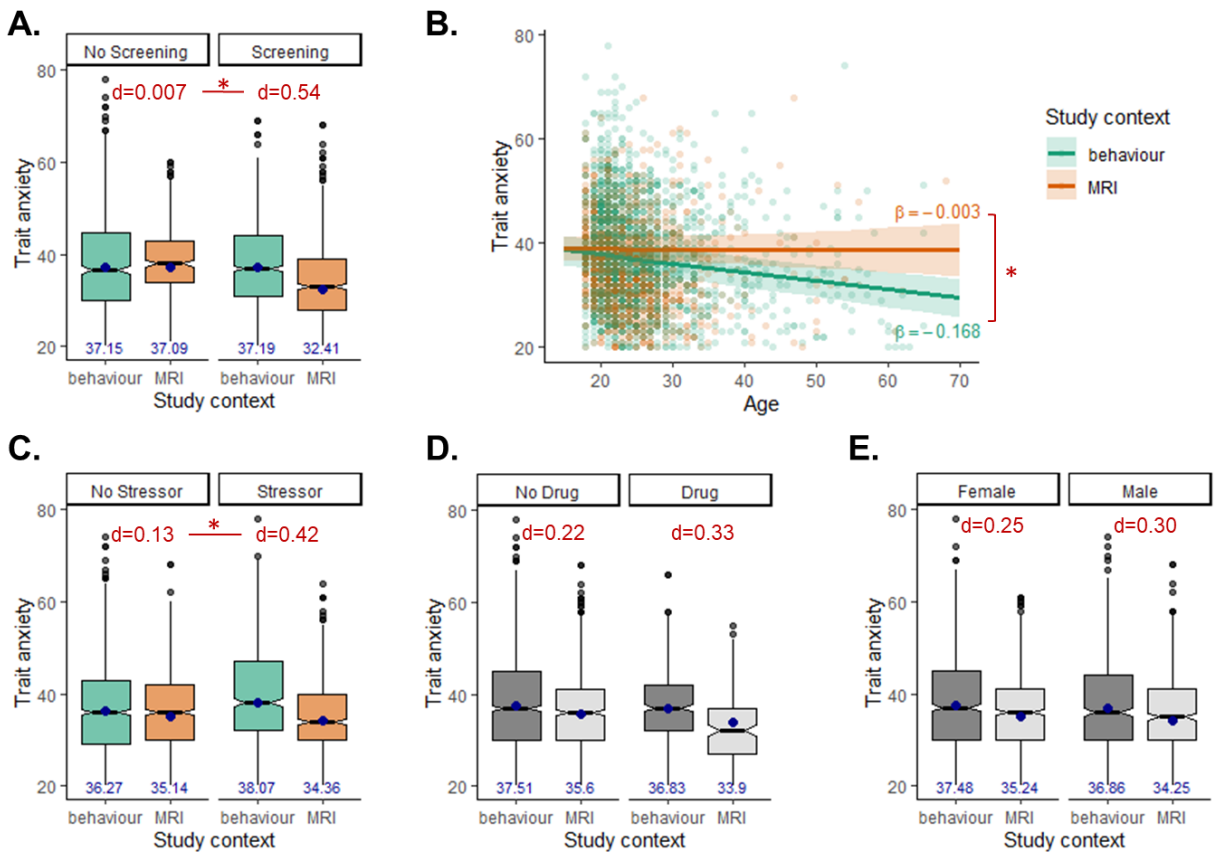

Figure 3. Moderating factors of the behaviour-MRI difference in trait anxiety (Model 2). Interaction effects with study context were added to the mixed effects models: (**A**) context-by-screening, (**B**) context-by-age, (**C**) context-by-stressor, (**D**) context-by-drug, and (**E**) context-by-gender. Effects of categorical factors (**A, C-E**) 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-MRI differences in marginal means. (**B**) Scatter plot of trait anxiety as a function of age (dots: raw data; lines: effect of age per condition predicted by the model). The interactions with screening, age and stressor (**A-C**) were found to be significant both in the mixed effects model and using Bayesian tests (* P<0.01 and B_{10}>3), whereas the interactions with drug and gender (**D-E**) were not.

Screening type	Site	Behaviour N	Behaviour Trait anxiety (± SD)	MRI N	MRI Trait anxiety (± SD)	Difference T	Difference P	Difference d
No screening	***all***	***728***	***38.68 (±11.08)***	***473***	***38.99 (±7.30)***	***-0.58***	***0.56***	***0.031***
	#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	***all***	***786***	***39.33 (±9.31)***	***525***	***34.36 (±7.78)***	***10.46***	***<0.001***	***0.57***
	#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	***all***	***462***	***33.90 (±8.09)***	***343***	***33.50 (±9.04)***	***0.64***	***0.52***	***0.047***
	#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, MRI). 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.

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=-3.117 ±0.42 (SE), χ^2=55.25, P<0.0001, $BF_{10}>10^{10}$). The effect size of the difference in marginal means was 0.355 (averaged over the levels of all other factors). There was also a significant main effect of psychiatric screening type (χ^2=86.50, P<0.0001, $BF_{10}>10^{16}$, **Fig 4A**), with higher trait anxiety for unscreened compared to fully screened individuals (estimate=2.877 ±0.56 (SE)), and for individuals screened by phone compared to those that were fully screened in person (estimate=6.265 ±0.67 (SE)). Re-running linear mixed model 2, including a random effect of site and interaction effects with study context, showed a significant interaction between study context and the type of screening procedure (χ^2=86.50, P<0.0001, $BF_{10}>10^4$). Mean trait anxiety scores collapsed across all sites (**Table 4**) showed that the interaction was driven by lower trait anxiety for MRI 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). This was also confirmed in the specific study sites that employed the same screening procedure for both MRI and behaviour study contexts (**Table 4**).

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 MRI 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 80th percentile is lower for MRI studies (45 vs 48), confirming the narrower distribution of trait anxiety scores in MRI 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 80th percentile (40.2 vs 48) were lower in MRI 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, MRI=30.40; 80th percentile: behaviour=41, MRI=40).

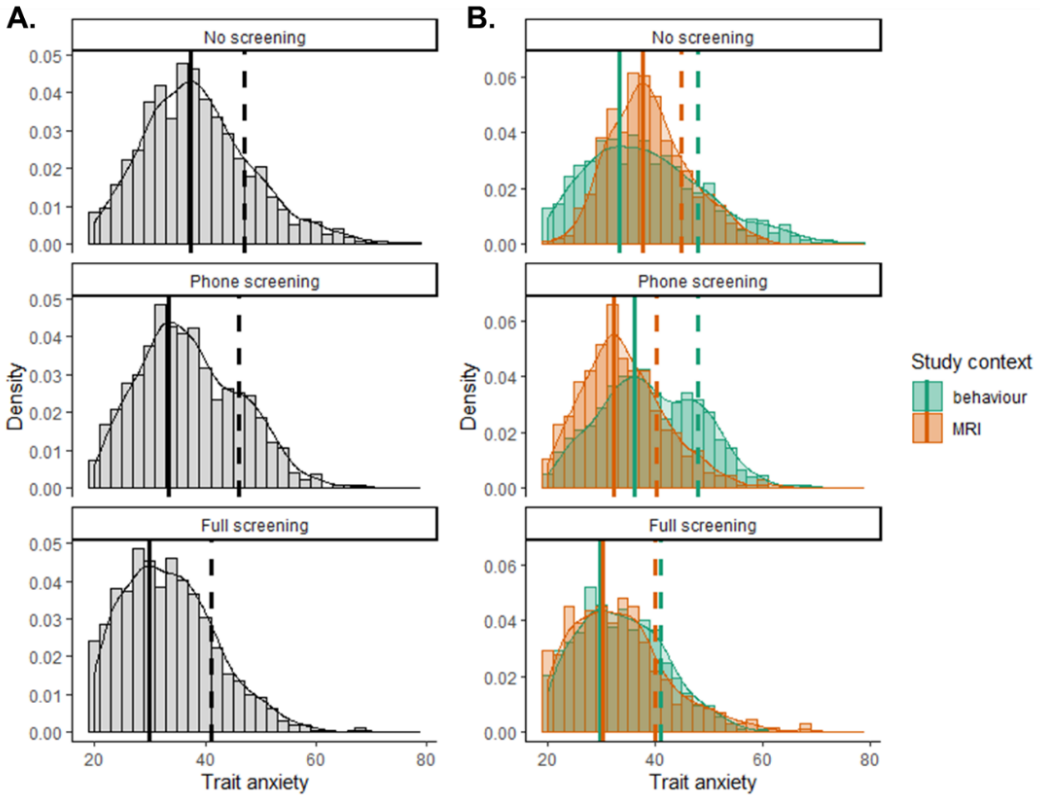

Figure 4. Distribution of trait anxiety scores split by screening procedure. Density plots of trait anxiety scores are shown (bin=2), separately for individuals who were not screened for psychiatric/affective disorders (top panels), screened by phone (middle panels) or fully screened with an in-person structured clinical interview (bottom panels). Solid lines show the mode of the distribution; dashed lines the 80th percentile. (**A**) Distribution for the entire population. (**B**) Separate distribution for behaviour and MRI study contexts.

Discussion

In this study, we provided substantial evidence, in a large-scale dataset across multiple sites, that individuals participating in MRI studies exhibit on average lower levels of trait anxiety than individuals participating in behavioural studies only. We showed that this effect is robust to controlling for multiple factors – namely age, gender, whether or not participants were screened for affective and psychiatric disorders, the type of screening procedure used, whether or not the study involved the presence of a stressor, and whether or not the study involved a pharmacological manipulation (drug administration). We also confirmed this effect using both mixed modelling approaches and Bayesian analyses. Interestingly, the difference in trait anxiety scores was also accompanied by a clear difference in their distribution across study contexts; specifically, a strikingly lower proportion of individuals had trait anxiety scores above 45 in MRI studies relative 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 MRI and behavioural study contexts. Specifically, when no or minimal screening was performed, the range of trait anxiety scores was narrower in MRI 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 MRI studies more often than behavioural studies, because of inherent differences in screening questions. For example, during recruitment of MRI 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 'lie' about or ignore 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 MRI studies, even if they meet all eligibility criteria. Undergoing MRI scanning has been found to be anxiogenic, because of claustrophobia, discomfort, and/or fear of learning about potential incidental findings[7–9]; 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 MRI studies may partly explain the bias[8,9], whether other specific components of anxiety play a role remains unclear. Factor analyses of the STAI[58–60] 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 MRI 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.

Our results also revealed that the difference in trait anxiety between MRI and behavioural studies was larger in studies involving a stressor. This interaction effect was observed without a main effect of stressor; that is, there was no overall difference in trait anxiety between stress and non-stress studies. Instead, it is only when the presence of a stress-induction procedure is combined with MRI that the behaviour-MRI difference in trait anxiety is the strongest. 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[34,61,62]. However, contrary to the literature suggesting both higher prevalence of anxiety disorders[63] and higher self-reported anxiety[34,64] in females than males, we did not find strong support for this hypothesis in our data. Our data indicated that trait anxiety was numerically higher for females, but the significance of that effect was not confirmed using Bayesian tests, consistent with the negligible effect size of that difference (Cohen's d=0.106).

While the large scale of the present dataset allowed us to ensure the robustness of the effects, with data obtained from multiple institutions and controlling for multiple potential confounds, we note two 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 controlled for in the analyses (age, gender, screening, stressor, drug, and study site) 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 MRI studies[65]. 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.

Overall, the finding of lower trait anxiety, as well as narrower distribution of trait anxiety scores, in MRI 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 MRI study (e.g [1,2]) or vice versa. Because of the narrower range of trait anxiety values in MRI 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. Taken together, these findings point towards a recommendation for cognitive neuroscience researchers who run both MRI and behavioural studies to measure individual differences in anxiety and carefully consider and mitigate potential sources of recruitment bias. Our result that distributions of trait anxiety scores between MRI and behavioural studies match almost perfectly when full in-person psychiatric screening interviews are conducted suggests that such screening procedure should be relied on when possible. This recommendation is particularly relevant for researchers running studies involving a stressor, in which the trait anxiety difference between behavioural and MRI components was found to be larger. Probing more deeply into individual reasons for participating in MRI studies and differences in screening procedures is thus necessary to ensure researchers can enforce a distribution of psychological and psychiatric profiles that is representative of the general population.

Methods

Procedure. Trait anxiety total scores, from the State Trait Anxiety Inventory[33] 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). 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 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, two mixed effects models were built (i) to examine the effect of study context (behaviour vs MRI) while controlling for 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. 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[66] in order to corroborate the effects obtained with mixed effects models. Bayesian ANCOVA 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, while controlling for all other fixed effects – this allowed determining the significance of main effects. Second, we compared pairs of models either including or not including an interaction effect of interest, while controlling for all fixed effects and all other interactions. 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_{10}), 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[67–70]. Evidence in favor 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 favor 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: effect of screening procedures. To examine the role of specific psychiatric screening procedures in modulating trait anxiety differences between MRI 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 80th percentile (**Fig. 4**) to characterize the distributions.

Data and code availability. Data and code are available on the following github repository: https://github.com/ccharpen/Trait_anxiety_MRI_BH

Acknowledgements

CJC, JOD, and LKP are supported by the Caltech Conte Center for the Neurobiology of Social Decision-Making (NIMH award: P50MH094258). CJC and JOD are 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 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

1. Bolton, S. & Robinson, O. J. The impact of threat of shock-induced anxiety on memory encoding and retrieval. *Learn. Mem.* **24**, 532–542 (2017).
2. Garibbo, M., Aylward, J. & Robinson, O. J. The impact of threat of shock-induced anxiety on the neural substrates of memory encoding and retrieval. *Soc. Cogn. Affect. Neurosci.* **14**, 1087–1096 (2019).
3. Hommel, B., Fischer, R., Colzato, L. S., van den Wildenberg, W. P. M. & Cellini, C. The effect of fMRI (noise) on cognitive control. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 290–301 (2012).
4. Skouras, S., Gray, M., Critchley, H. & Koelsch, S. fMRI scanner noise interaction with affective neural processes. *PLoS One* **8**, 1–13 (2013).
5. Kobald, S. O., Getzmann, S., Beste, C. & Wascher, E. The impact of simulated MRI scanner background noise on visual attention processes as measured by the EEG. *Sci. Rep.* **6**, 1–10 (2016).
6. Gossett, E. W. *et al.* Anticipatory stress associated with functional magnetic resonance imaging: Implications for psychosocial stress research. *Int. J. Psychophysiol.* **125**, 35–41 (2018).
7. Meléndez, J. C. & McCrank, E. Anxiety-Related Reactions Associated With Magnetic Resonance Imaging Examinations. *JAMA* **270**, 745–747 (1993).
8. Katz, R. C., Wilson, L. & Frazer, N. Anxiety and its determinants in patients undergoing Magnetic Resonance Imaging. *J. Behav. Ther. Exp. Psychiatry* **25**, 131–134 (1994).
9. Murphy, K. J. & Brunberg, J. A. Adult claustrophobia, anxiety and sedation in MRI. *Magn. Reson. Imaging* **15**, 51–54 (1997).
10. Beddington, J. *et al.* The mental wealth of nations. *Nature* **455**, 1057–1059 (2008).
11. Robinson, O. J., Pike, A. C., Cornwell, B. & Grillon, C. The translational neural circuitry of

anxiety. *J. Neurol. Neurosurg. Psychiatry* **90**, 1353–1360 (2019).
12. Grillon, C., Robinson, O. J., Cornwell, B. & Ernst, M. Modeling anxiety in healthy humans: a key intermediate bridge between basic and clinical sciences. *Neuropsychopharmacology* **0**, 1–12 (2019).
13. Robinson, O. J., Vytal, K., Cornwell, B. R. & Grillon, C. The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Front Hum Neurosci* **7**, 203 (2013).
14. Engel-Yeger, B. & Dunn, W. The relationship between sensory processing difficulties and anxiety level of healthy adults. *Br. J. Occup. Ther.* **74**, 210–216 (2011).
15. Grillon, C. Startle reactivity and anxiety disorders: aversive conditioning, context, and neurobiology. *Biol. Psychiatry* **52**, 958–975 (2002).
16. Poli, E. & Angrilli, A. Greater general startle reflex is associated with greater anxiety levels: A correlational study on 111 young women. *Front. Behav. Neurosci.* **9**, 1–6 (2015).
17. Bar-Haim, Y., Lamy, D. & Pergamin, L. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychol. Bull.* **133**, 1–24 (2007).
18. Cisler, J. M. & Koster, E. H. W. Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clin. Psychol. Rev.* **30**, 203–216 (2010).
19. Etkin, A., Prater, K. E., Hoeft, F., Menon, V. & Schatzberg, A. F. Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *Am J Psychiatry* **167**, 545–554 (2010).
20. Farmer, A. S. & Kashdan, T. B. Social anxiety and emotion regulation in daily life: spillover effects on positive and negative social events. *Cogn Behav Ther* **41**, 152–62 (2012).
21. Bishop, S. J. Trait anxiety and impoverished prefrontal control of attention. *Nat Neurosci* **12**, 92–98 (2009).
22. Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A. & Davidson, R. J. Anxiety selectively disrupts visuospatial working memory. *Emotion* **6**, 40–61 (2006).
23. Yao, N., Chen, S. & Qian, M. Trait anxiety is associated with a decreased visual working memory capacity for faces. *Psychiatry Res.* **270**, 474–482 (2018).
24. Browning, M., Behrens, T. E., Jocham, G., Reilly, J. X. O. & Bishop, S. J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* **18**, 590–596 (2015).
25. Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P. & Robinson, O. J. Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. *Biol. Psychiatry* **82**, 532–539 (2017).
26. Maner, J. K. *et al.* Dispositional anxiety and risk-avoidant decision-making. *Pers. Individ. Dif.* **42**, 665–675 (2007).
27. Clark, L. *et al.* Risk-avoidant decision making increased by threat of electric shock. *Psychophysiology* **49**, 1436–1443 (2012).
28. Charpentier, C. J., Aylward, J., Roiser, J. P. & Robinson, O. J. Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biol. Psychiatry* **81**, 1014–1022 (2017).
29. Etkin, A. *et al.* Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron* **44**, 1043–1055 (2004).
30. Stein, M. B., Simmons, A. N., Feinstein, J. S. & Paulus, M. P. Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. *Am J Psychiatry* **164**, 318–327 (2007).
31. Robinson, O. J. *et al.* 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 (2014).

32. Carlisi, C. O. & Robinson, O. J. The role of prefrontal–subcortical circuitry in negative bias in anxiety: Translational, developmental and treatment perspectives. *Brain Neurosci. Adv.* **2**, 239821281877422 (2018).
33. Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R. & Jacobs, A. G. *Manual for the State-Trait Anxiety Inventory*. (Consulting Psychologists Press, 1983).
34. Knight, R. G., Waal-Manning, H. J. & Spears, G. F. Some norms and reliability data for the State-Trait Anxiety Inventory and the Zung Self-Rating Depression scale. *Br. J. Clin. Psychol.* **22**, 245–249 (1983).
35. Julian, L. J. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). *Arthritis Care Res.* **63**, 467–472 (2011).
36. Fisher, P. L. & Durham, R. C. Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the STAI-T across outcome studies since 1990. *Psychol. Med.* **29**, 1425–1434 (1999).
37. Kennedy, B. L., Schwab, J. J., Morris, R. L. & Beldia, G. Assessment of State and Trait Anxiety in Subjects with Anxiety and Depressive Disorders. *Psychiatr. Q.* **72**, 263–276 (2001).
38. Charpentier, C. J., De Martino, B., Sim, A. L., Sharot, T. & Roiser, J. P. Emotion-induced loss aversion and striatal-amygdala coupling in low-anxious individuals. *Soc. Cogn. Affect. Neurosci.* **11**, 569–579 (2016).
39. Charpentier, C. J., De Neve, J.-E., Li, X., Roiser, J. P. & Sharot, T. Models of affective decision making: how do feelings predict choice? *Psychol. Sci.* **27**, 763–775 (2016).
40. Charpentier, C. J., Bromberg-Martin, E. S. & Sharot, T. Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proc. Natl. Acad. Sci.* **115**, E7255–E7264 (2018).
41. Kausche, F. M. & Schwabe, L. Blocking under stress: Sustained attention to stimuli without predictive value? *Neurobiol. Learn. Mem.* **168**, 107158 (2020).
42. Kalbe, F. & Schwabe, L. Beyond arousal: Prediction error related to aversive events promotes episodic memory formation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **46**, 234–246 (2020).
43. Balderston, N. L. *et al.* Mechanistic link between right prefrontal cortical activity and anxious arousal revealed using transcranial magnetic stimulation in healthy subjects. *Neuropsychopharmacology* **45**, 694–702 (2020).
44. Balderston, N. L. *et al.* Low-frequency parietal repetitive transcranial magnetic stimulation reduces fear and anxiety. *Transl. Psychiatry* **10**, 68 (2020).
45. Lago, T. R. *et al.* Exercise decreases defensive responses to unpredictable, but not predictable, threat. *Depress. Anxiety* **35**, 868–875 (2018).
46. Murphy, S. E., Wright, L. C., Browning, M., Cowen, P. J. & Harmer, C. J. A role for 5-HT4 receptors in human learning and memory. *Psychol. Med.* 1–9 doi:DOI: 10.1017/S0033291719002836
47. Huys, Q. J. M. *et al.* Interplay of approximate planning strategies. *Proc Natl Acad Sci U S A* **112**, 3098–3103 (2015).
48. Lally, N. *et al.* The Neural Basis of Aversive Pavlovian Guidance during Planning. *J. Neurosci.* **37**, 10215 LP – 10229 (2017).
49. Peters, S. E. *et al.* Cognitive bias modification for facial interpretation: A randomized controlled trial of transfer to self-report and cognitive measures in a healthy sample. *R. Soc. Open Sci.* **4**, (2017).
50. Faulkner, P., Selvaraj, S., Pine, A., Howes, O. D. & Roiser, J. P. The relationship between reward and punishment processing and the 5-HT1A receptor as shown by PET.

Psychopharmacology (Berl). **231**, 2579–2586 (2014).
51. Ly, V., Cools, R. & Roelofs, K. Aversive disinhibition of behavior and striatal signaling in social avoidance. *Soc. Cogn. Affect. Neurosci.* **9**, 1530–1536 (2013).
52. Zerbes, G. & Schwabe, L. Across time and space: spatial-temporal binding under stress. *Learn. Mem.* **26**, 473–484 (2019).
53. Zerbes, G., Kausche, F. M., Müller, J. C., Wiedemann, K. & Schwabe, L. Glucocorticoids, Noradrenergic Arousal, and the Control of Memory Retrieval. *J. Cogn. Neurosci.* **31**, 288–298 (2018).
54. Wanke, N. & Schwabe, L. Subjective Uncontrollability over Aversive Events Reduces Working Memory Performance and Related Large-Scale Network Interactions. *Cereb. Cortex* (2019). doi:10.1093/cercor/bhz298
55. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
56. Rouder, J. & Morey, R. Default Bayes factors for model selection in regression. *Multivar Behav Res* **47**, 877–903 (2012).
57. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *J. Math. Psychol.* **56**, 356–374 (2012).
58. Bieling, P. J., Antony, M. M. & Swinson, R. P. The state-trait anxiety inventory, trait version: Structure and content re-examined. *Behav. Res. Ther.* (1998). doi:10.1016/S0005-7967(98)00023-0
59. Vigneau, F. & Cormier, S. The Factor Structure of the State-Trait Anxiety Inventory: An Alternative View. *J. Pers. Assess.* **90**, 280–285 (2008).
60. Wang, T. *et al.* The Factorial Structure of Trait Anxiety and Its Mediating Effect Between Mindfulness and Depression. *Front. psychiatry* **9**, 514 (2018).
61. Nakazato, K. & Shimonaka, Y. The Japanese State-Trait Anxiety Inventory: Age and Sex Differences. *Percept. Mot. Skills* **69**, 611–617 (1989).
62. Regier, D. A., Narrow, W. E. & Rae, D. S. The epidemiology of anxiety disorders: The epidemiologic catchment area (ECA) experience. *J. Psychiatr. Res.* **24**, 3–14 (1990).
63. McLean, C. P., Asnaani, A., Litz, B. T. & Hofmann, S. G. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* **45**, 1027–1035 (2011).
64. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).
65. Leach, L. S., Christensen, H., Mackinnon, A. J., Windsor, T. D. & Butterworth, P. Gender differences in depression and anxiety across the adult lifespan: the role of psychosocial mediators. *Soc. Psychiatry Psychiatr. Epidemiol.* **43**, 983–998 (2008).
66. Love, J. *et al.* JASP (Version 0.7.1) [Computer Software]. (2015).
67. Lee, M. & Wagenmakers, E.-J. *Bayesian cognitive modeling: a practical course.* (Cambridge University Press, 2014).
68. Jarosz, A. F. & Wiley, J. What are the odds? A practical guide to computing and reporting Bayes Factors. *J. Probl. Solving* **7**, 2–9 (2014).
69. Quintana, D. S. & Williams, D. R. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry* **18**, 178 (2018).
70. Kass, R. E. & Raftery, A. E. Bayes Factors. *J Am Stat Assoc* **90**, 773–795 (1995).

Supplementary Information

Study context	Gender	Psychiatric screening	Stressor	Drug administration	Mean trait anxiety	SD	N
MRI	F	0	0	0	39.09	8.30	151
				1	-	-	0
			1	0	38.49	5.96	90
				1	-	-	0
		1	0	0	35.46	9.02	101
				1	32.67	7.28	75
			1	0	35.03	8.24	241
				1	32.48	7.48	29
	M	0	0	0	39.23	6.82	152
				1	-	-	0
			1	0	38.89	7.65	80
				1	-	-	0
		1	0	0	35.14	10.17	116
				1	33.31	7.69	71
			1	0	32.85	7.31	224
				1	28.55	7.12	11
behaviour	F	0	0	0	39.04	11.09	367
				1	-	-	0
			1	0	39.72	11.63	85
				1	-	-	0
		1	0	0	33.89	8.14	198
				1	38.30	7.97	125
			1	0	42.03	10.25	204
				1	36.61	8.42	117
	M	0	0	0	37.93	11.07	207
				1	-	-	0
			1	0	37.70	10.31	69
				1	-	-	0
		1	0	0	34.88	9.84	156
				1	37.40	6.76	149
			1	0	38.22	10.15	180
				1	36.34	8.14	119

Table S1. Trait anxiety descriptive statistics, collapsed across study sites (N=3317). Mean trait anxiety, standard deviation, and sample size are reported, broken down by each of the five categorical factors collected in the data.

	Behaviour		MRI	
	N	**Screening procedure**	**N**	**Screening procedure**
Site #1	100	**None:** study flyer specifying no psychiatric disorder	155	**Full:** MINI
Site #2	102	**None:** no exclusion criteria for psychiatric or neurological disorders	0	-
Site #3	425	**Phone:** smoking, medication intake, drug use, history of neurological/ psychiatric diagnosis, history of drug abuse.	465	**Phone:** same interview as Behaviour
Site #4	71	**Phone:** BMI, medication use, alcohol use, smoking, drug use, severe illness in past year, diabetes, psychiatric problems in past year requiring therapy, follow-up MDD symptoms	0	-
Site #5	55	**None:** self-report of meeting eligibility criteria stated on information letter – no neurological, cardiovascular diseases, psychiatric disorders, regular use of medication or marijuana, use of psychotropic drugs	45	**None:** same as Behaviour + claustrophobia & heavy smoking
Site #6	27	**None:** self-report of meeting eligibility requirements – no psychoactive drugs, no psychological conditions, no neurological condition	413	**None:** same as Behaviour
Site #7	33	**Full:** SCID	61	**Full:** SCID
Site #8	188	**Full:** SCID	40	**Full:** SCID
	30	**Phone:** exclude past history of psychiatric or neurological disorder		
	168	**None:** self-report of no use of psychotropic medication, no diagnosis of current Axis I disorder, BMI between 18 and 30	15	**None:** same as Behaviour
Site #9	241	**Full:** MINI	87	**Full:** MINI
	260	**Phone:** past/present diagnosis of depression, stress-related problems, bipolar disorder, ADHD, eating disorder, OCD, trichotillomania, learning disability, alcohol & drug abuse, medication for psychiatric disorder	60	**Phone:** same interview as Behaviour + claustrophobia
	276	**None:** self-report of meeting eligibility criteria specified on study advert – no past or present psychiatric or neurological disorder, alcohol use (past 24hrs) or cannabis (past week).		

Table S2. Details of screening procedures across study sites and study contexts. The number of individual data points for each site and each study context are reported, as well as details about the screening procedure and eligibility criteria. SCID: Structured Clinical Interview for DSM-5 Axis I disorders. MINI: Mini-International Neuropsychiatric Interview. BMI: Body-Mass Index. MDD: Major Depressive Disorder. ADHD: Attention Deficit Hyperactivity Disorder. OCD: Obsessive Compulsive Disorder.