Appendix I. *Tests for Assumptions of Parametric Analyses*

Please evaluate the results of these analyses in conjunction with Appendix H (Data Analysis Script). For the ANOVAs and ANCOVAs, data were checked to ensure that the dependent variables and covariate were measured on a continuous scale, and that the independent variable consisted of two categorical and independent groups. The variables were transformed into numeric and factor variables, respectively, to ensure these assumptions would hold.

Data were then checked to ensure that assumptions of normally distributed outcome variables, no extreme outliers, homogeneity of variance, sphericity, homogeneity of variance-covariance matrices were met for the ANOVAs. Additional assumptions required for parametric testing utilising the ANCOVA were also tested: independence of the covariate and IVs, homogeneity of regression slopes, and linearity between the covariate and DVs at each level of IV groupings.

Visual inspection of Q-Q plots, as well as computation of the Shapiro-Wilk test, indicated that fixation count was not normally distributed at all levels of IU, Stimulus, and Time (*p*s < .05), with the exception of high IU late extinction CS- (*p* = 391) and low IU early extinction CS+ (*p* = 228). Log-transformed fixation duration was normally distributed at all levels of IU, Stimulus, and Time (*p*s > .05) with the exception of low IU acquisition CS+ (*p* = .039), high IU late extinction CS- (*p* = .024), and low IU early extinction CS- (*p* = .004). Saccade amplitude was not normally distributed at any levels of the IVs (all *p*s < .05). Though data violated the assumption of normality, according to the central limit theorem (Lumley et al., 2002), the large sample size (*n* = 139) was considered to be normally distributed.

Outliers were identified using the *identify_outliers* function from the *rstatix* package in R (version 0.7.0, Kassambara, 2021), which uses boxplot methods to detect outliers. Values that exceed the third quartile by 3 x IQR or below the first quartile by 3 x IQR are considered to be extreme outliers. This identified a total of four extreme outliers: two for fixation count in extinction and two for saccade amplitude in extinction. This assumption was therefore violated.

There was homogeneity of variances for both IU groups in all variables as assessed by Levene's test (all *p*s > .05) with the exception of log-transformed fixation duration at the level of early extinction and CS- [$F(1,137) = 8.18$, $p = .005$], log-transformed fixation duration at the level of late extinction and CS+ [$F(1,137) = 7.14$, $p = .008$], and log-transformed fixation duration at the level of late extinction and CS- [$F(1,137) = 7.74$, $p = .006$]. However, in large samples, Levene's test can yield significant results even when group variances do not largely differ (Field et al., 2012).

The assumption of sphericity was not applicable in this case, as there were no within-subjects factors with > 3 levels.

Covariance matrices were equal across cells formed by the between-subjects factor (IU) group for all dependent variables (*p*s > .05), as assessed by Box's test of equality of covariance

matrices, with the exception of log-transformed fixation duration in extinction ($p < .001$). However, sample sizes were equal between IU groups (High IU: $n = 68$; Low IU: $n = 71$). Therefore, the violation of this assumption was ignored in this case.

Independence of the covariate and grouping conditions (IU, Stimulus, Time) was assessed using independent-samples *t*-tests, with each grouping condition entered as the independent variable and mean-centred STICSA scores (the covariate within the ANCOVA) entered as the dependent variable. Mean-centred STICSA was independent of Stimulus and Time across acquisition and extinction (all *p*s > .05) . However, STICSA was not independent of IU across acquisition and extinction (all *p*s < .05). This was not surprising or concerning as IU is considered to be a lower order factor of anxiety (Carleton, 2016), and so the two constructs are not likely to be independent.

To test the assumption of invariance of regression slopes between conditions, an interaction term of conditions (IU, Stimulus, Time) and mean centred STICSA totals was computed, and inspected visually using scatterplots with regression lines. There was homogeneity of regression slopes as the interaction term was not statistically significant (*p*s > .05) for all combinations with the exception of saccade amplitude in extinction, where mean-centred STICSA and IU group shared a significant interaction ($p = .005$).

The assumption of linearity between the covariate and outcome variables was tested through visual inspection of grouped scatterplots. There were linear relationships between mean-centred STICSA and fixation count, fixation duration, and saccade amplitude at all levels of IU, Stimulus, and Time. This assumption was therefore met.

Overall, the majority of assumptions were met for the majority of DVs (fixation count, fixation duration, and saccade amplitude) at levels of the IVs (IU, Stimulus and Time). It was therefore concluded that it was appropriate to run parametric tests to compute inferential statistics from the sample.