Supplementary Methods

S1. Selection and Matching of Image Pairs

In our selection of stimuli, we excluded images that were i) low resolution ii) black and white iii) had a political connotation iv) Involved famous people or events v) had sexually explicit content. Social images generally included pictures of happy individuals including couples, families, babies and children. Nonsocial images included rewarding content that did not include people, including food, scenery and money.

A pilot sample of 23 participants (12 females) gave valence and arousal ratings for each image via a standard self-assessment manikin procedure [1] . Observers made ratings on a 9 point likert scale, where a 5 rating conveys a neutral valence/arousal, and 6 or higher conveys a rewarding image or one that makes them feel 'excited/jittery'. All participants completed the valence ratings, but 2 males did not complete the arousal ratings.

The Koch toolbox [2] was used to calculate image saliency for each pair of images. The Koch Toolbox produces a composite measure of image saliency including features such as colour contrast and edge orientation calculated for each image. The global root mean squared (RMS) and local RMS contrast for each image were also computed as described in [3].

To ensure that the image pairs presented on any given trial during the experiment were matched as closely as possible in dimensions other than their sociality, we chose the image pairings that minimised the difference between all affective (valence, arousal) and salience (RMS contrast - Global, RMS contrast - Local, Koch Salience) metrics. One sample t tests revealed that there were no significant mean differences between social and nonsocial image categories in any metric (Figure S1).

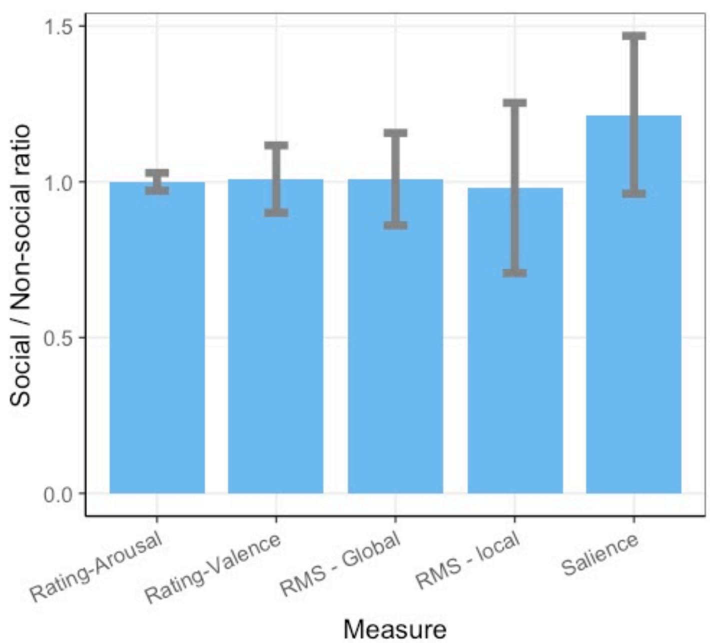

Figure S1. Comparisons of the social and nonsocial stimulus pairs including: arousal ratings; valence ratings; global root mean square (RMS) contrast; local RMS contrast; and stimulus saliency using Koch Toolbox [2]. The error bars are 95% confidence intervals

S2. Cross Validation Procedure

To test the generalisation performance of our models and avoid overfitting, we performed a standard 'leave one out' analysis. We first created a *training set*, consisting of *N*-1 observers and a *test set*, consisting of just the 'left out' observer. A model was fit to the *training set* and then the parameters of this model were used to predict the empirical data for the *test set*. We then calculated the sum of squared residuals (SSR) to index the performance of the model in predicting the *test set*. We repeated this process *N* times, each time leaving out a different observer. To summarise the performance of each model into a single metric (LOO_{xval}), we summed the SSR across the *N* test sets.

S3. Dataset 1: Models. Full data.

Model	Parameters	AIC	LOO_{xval}
1	Intercept only	-5169.9	143.86
2	AOI	-6176.4	121.99
3	AOI x linear	-6771.3	120.36
4	AOI x quadratic	-6266.1	120.72
5	AOI x cubic	-6186.2	121.82
6	**AOI x (linear + quadratic)**	**-6365.5***	**118.56***
7	AOI x (linear + cubic)	-6263.8	120.38
8	AOI x (quadratic + cubic)	-6243.1	120.75
9	AOI x (linear + quadratic + cubic)	-6353.8	118.71

Asterisk indicates the best performing model.

S4. Cluster Definition and Permutation Analysis.

Our analysis revealed several time bins wherein an effect of EQ was detected. The rationale for grouping any such adjacent time bins into clusters is that one such time bin in isolation is more likely to reflect 'noise', whereas several adjacent time bins wherein the null hypothesis is rejected are more likely to reflect 'signal'. Therefore, 'clusters' of contiguous time bins are defined so that the size of the cluster can be incorporated into the correction for multiple comparisons and protect against type 1 errors (see below). Defining clusters in this way implicitly penalizes isolated time bins (more likely to be noise) relative to several adjacent time bins (more likely to be signal).

Permutation analysis, described in detail in [4] controls for familywise error rate associated with multiple tests, whilst also taking into account the statistical dependencies in time-series data. Briefly, an initial test statistic is obtained for each time bin. Contiguous series of significant time bins are then defined (clusters). Within each cluster, the sum of statistics is obtained. Next, this process is repeated on 1000 randomly shuffled datasets, to obtain a null distribution of summed statistics. The *p* values reported in the main text therefore reflect the proportion of summed statistics in this null distribution that exceed that obtained from the empirical cluster.

S5. Dataset 1: Models involving EQ.

Model	Parameters	AIC	*LOO* $_{xval}$
1	AOI x linear + quadratic	-6304.1	117.29
2	*Reduced Interactive* AOI x (linear + quadratic) +(EQ*AOI)	-6637.5	111.47
3	*Fully Interactive* **AOI x EQ x (linear + quadratic)**	**-6702.2***	**110.79***

S6. Dataset 2: Models. Full data.

Model	Parameters	AIC	*LOO* $_{xval}$
1	Intercept only	-7590.5	167.40
2	AOI	-9498.6	132.38
3	AOI x linear	-9587.2	131.08
4	AOI x quadratic	-9539.1	131.84
5	AOI x cubic	-9490.7	132.57
6	**AOI x (linear + quadratic)**	**-9639.3***	**130.34***
7	AOI x (linear + cubic)	-9578.8	131.24
8	AOI x (quadratic + cubic)	-9530.4	132.05
9	AOI x (linear + quadratic + cubic)	-9632.5	130.47

S7. Dataset 2: Models involving EQ.

Model	Parameters	AIC	*LOO* $_{xval}$
1	AOI x linear + quadratic	-8374.5	121.57
2	*Reduced Interactive* AOI x (linear + quadratic) +(EQ*AOI)	-8706.5	117.55
3	*Fully Interactive* **AOI x EQ x (linear + quadratic)**	**-8783.6***	**116.83***

Data Availability

Both datasets supporting this manuscript can be found at the following URL:

https://figshare.com/articles/Eyetracking_2018_Dataset_1_and_2_/6876455

References

1. Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry. 1994 Mar 1;25(1):49–59.

2. Walther D, Koch C. Modeling attention to salient proto-objects. Neural Netw. 2006 Nov;19(9):1395–407.

3. Bex PJ, Makous W. Spatial frequency, phase, and the contrast of natural images. J Opt Soc Am. 2002;19(6):1096.

4. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007 Aug 15;164(1):177–90.