Supporting Information (SI Appendix) for

How cognitive and reactive fear circuits optimize escape decisions in humans

S. Qi[1,2]*, D. Hassabis[3], Jiayin Sun[2,6], F. Guo[4], N. Daw[5], D. Mobbs[1,2]*,

correspondence to:

DM: dmobbs@caltech.edu

SQ: sqi@caltech.edu

Fig. S1.
ANOVA for A) FID, B) Reward and C) Escapability rating.

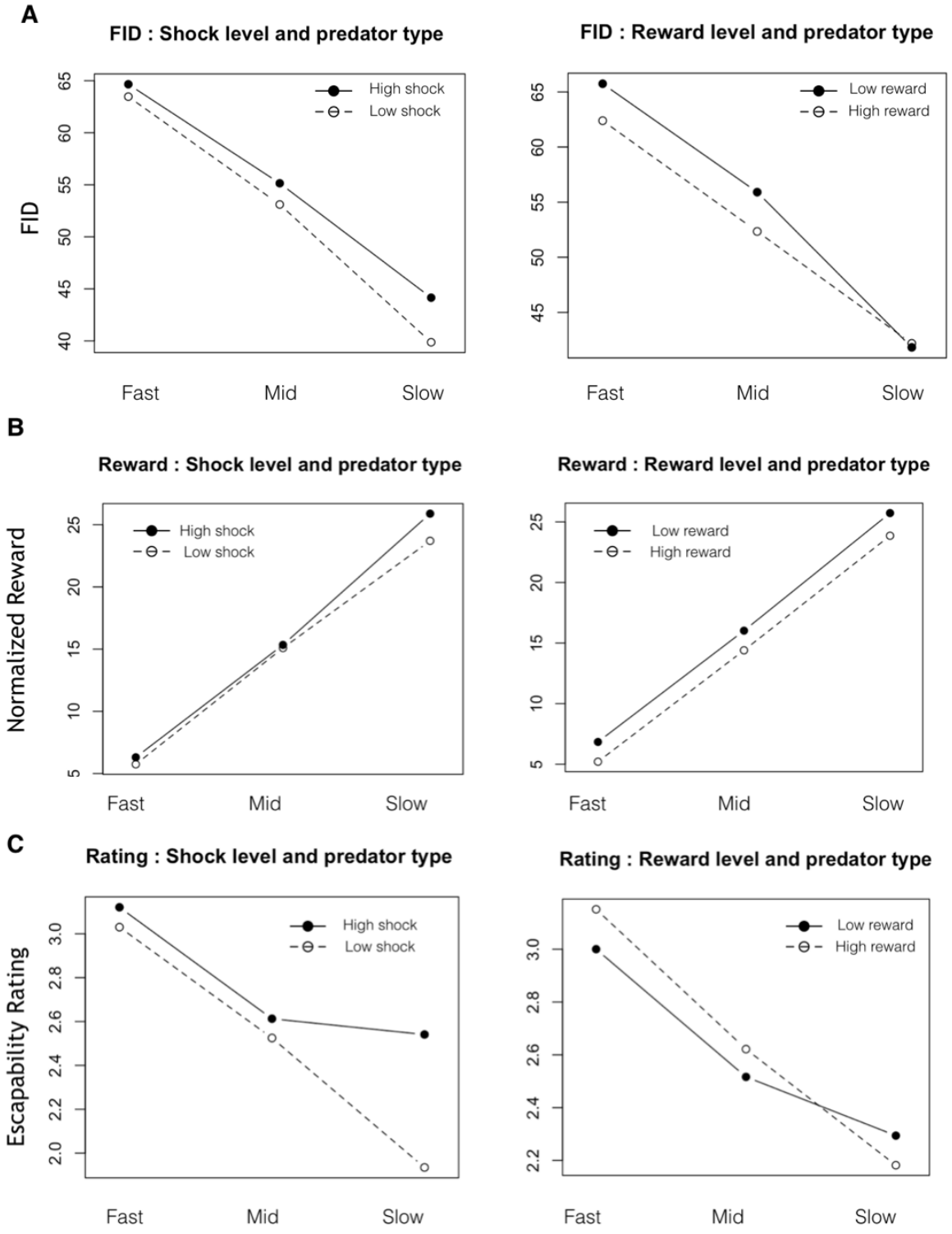

Fig. S2

Activated regions for 1st level parametric modulation with reward and punishment sensitivity parameters in the Bayesian decision making model.

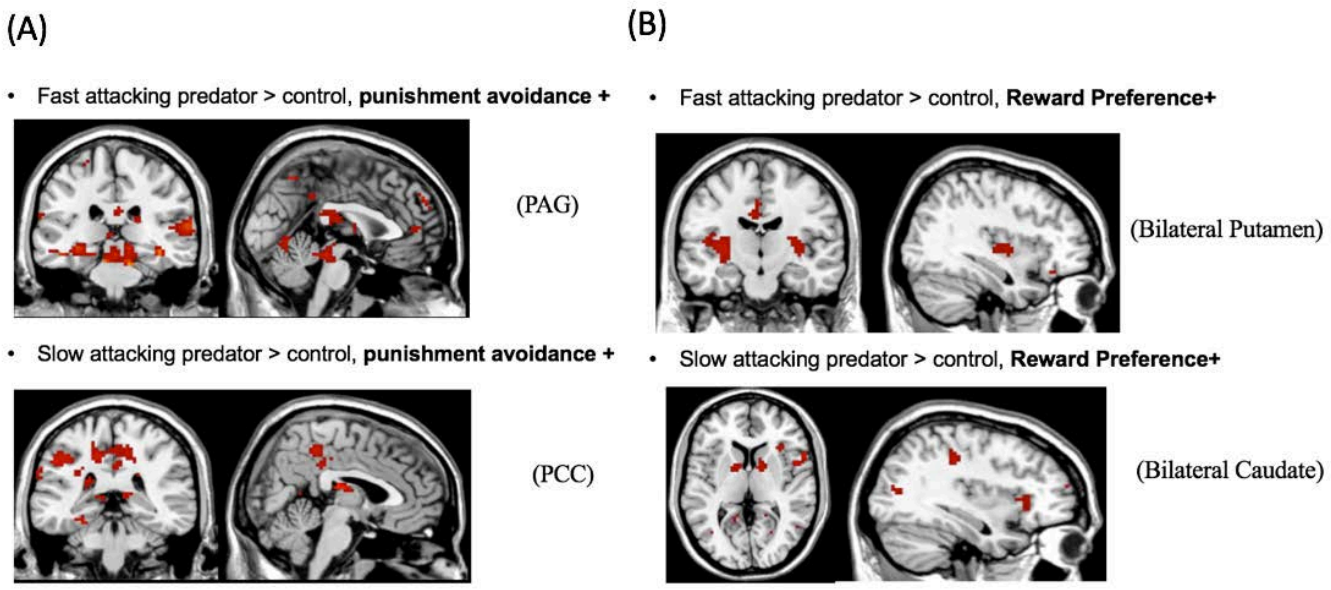

Fig. S3

(A) Relationship between FID and the chance of escape for each predator type. Chance of escape increases with FID, but with different growth patterns in every predator type. (B) Bayesian ideal observer estimates of predator AD, based on the unknown-mean-known-variance Gaussian ideal learner model, as a function of experience in the task. Color 1 (blue), 2 (yellow), 3 (red) corresponds to fast, mid and slow predators, respectively. The graph shows the observer's 95% credible interval almost always contains the true mean, indicating an appropriate modeling of uncertainty

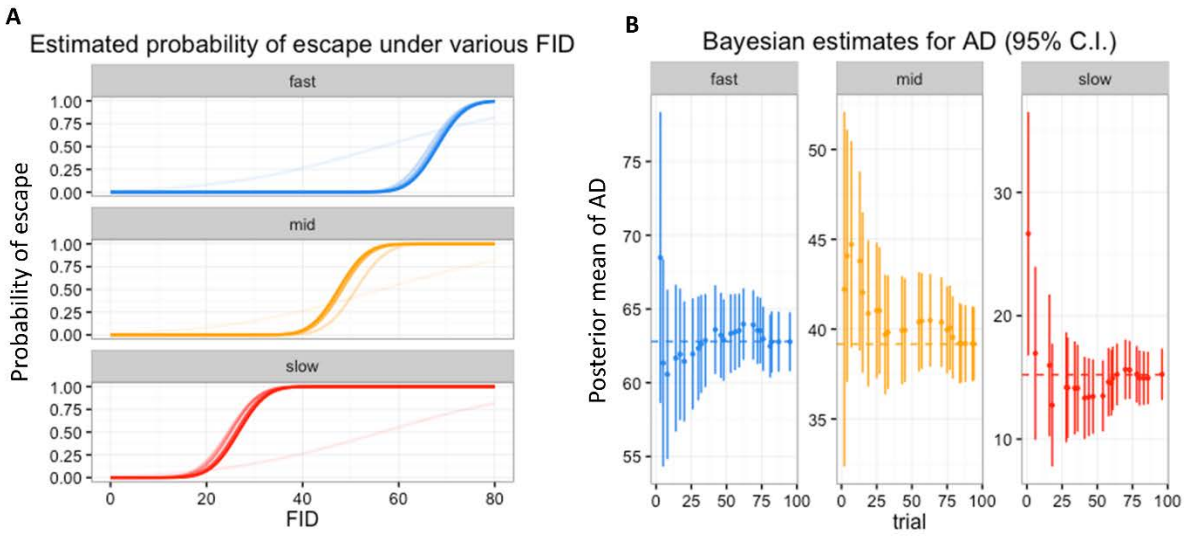

Fig. S4

(A)Estimated coefficients for each subject for the second scanning session, along with 95% confidence intervals. X axis represents the pain coefficient $\boldsymbol{\beta_1}$ in the utility function, and Y axis represents the monetary reward coefficient $\boldsymbol{\beta_2}$. For a rational player, β_2 should be positive (seeking money) and $\boldsymbol{\beta_1}$ should be negative (avoiding shock). (B) Model fits to observed FIDs for the second scanning session. X axis represents trial numbers, and Y axis represents FID. Ideal FID choices predicted by the ideal Bayesian observer (lines), subjects' actual FID choice (dots). Note that the colors here represent predator types (blue = fast attacking; red = slow attacking), not actual colors of the predators.

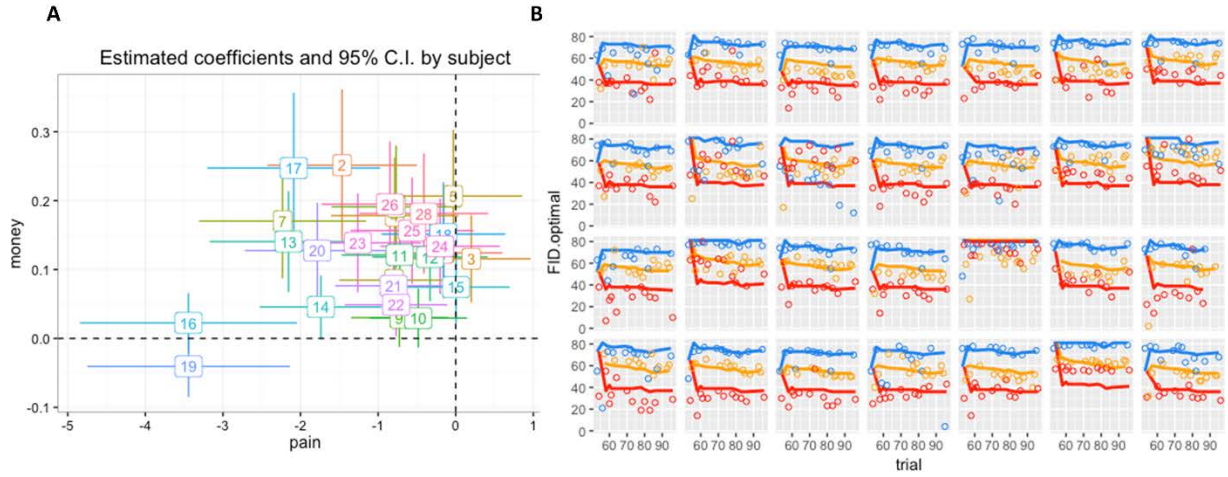

Table S1. Activation Table for Contrast [Far Predator > Control] (Whole Brain)

*** P<0.05, FDR corrected**

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Far Predator > Control						
Middle temporal Gyrus	R	461	7.82	48	-66	6
Middle temporal Gyrus	L	155	6.19	-42	-69	6
Precuneus	R	302	5.99	6	-48	48
Precentral Gyrus	L	81	4.52	-42	-9	48
Superior Temporal Gyrus	L	94	5.11	-57	-42	21
Insula	L	199	4.96	-33	21	3
Insula	R	285	6.72	37	20	-1
Mid Cingulate Gyrus	L	71	4.36	-12	-21	39
Midbrain		244	6.51	3	-28	-12
Supplementary Motor Area	R	357	6.28	17	5	60
Caudate	R	29	4.67	8	10	-5

Table S2. Activation Table for Contrast [Close Predator > Control] (Whole Brain)

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Close Predator > Control						
Middle temporal Gyrus	L	724	7.89	-54	-18	-9
Parahippocampal Gyrus	R	47	5.51	24	-12	-24
Supplementary Motor Area	R	213	5.31	3	-24	60
Middle Frontal Gyrus	R	26	3.91	33	36	-18
Medial Prefrontal Cortex	L	77	5.46	-3	56	-9
Insula	L	132	4.24	-33	-18	16
Insula	R	187	4.30	39	-11	15
Posterior Cingulate Cortex	R	423	5.65	6	-48	27

Table S3. Activation Table for Trial-by-trial parametric analysis with FID

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
FID parametric modulation + : Far Predator						
Midbrain	L	280	7.80	-12	27	-3
Hippocampus	R	31	7.03	18	-13	-13
Thalamus	L	90	6.87	-11	-26	-1
Insula	L	63	5.23	-38	5	-12
MCC	L	105	6.94	-3	2	42
FID parametric modulation - : Far Predator						
Middle Temporal Gyrus	L	51	3.63	-51	-15	-15
Middle Temporal Gyrus	R	40	3.45	51	-27	-15
Middle Occipital Gyrus	R	47	4.07	30	-90	15
FID parametric modulation + : Close Predator						
Inferior Frontal Gyrus	R	64	4.79	36	21	-6
Middle Frontal Gyrus	R	93	5.76	39	60	3
MCC	L	79	4.73	-1	33	36
Cerebrum	R	66	7.88	24	-9	36
FID parametric modulation - : Close Predator						
Superior Temporal Gyrus	L	29	5.56	-57	-51	12
Inferior Frontal Gyrus	L	43	6.46	-54	27	18
PCC	L	33	5.19	-9	-51	27

Table S4. Activation Table for Trial-by-trial parametric analysis with Escapability

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Escapability parametric modulation + : Fast Predator						
Middle Temporal Gyrus	L	104	7.52	42	-51	-21
Midbrain	R	221	9.20	9	-30	-15
Thalamus	R	75	10.42	12	-7	3
Thalamus	L	42	9.54	-5	-12	4
Insula	L	61	9.88	-41	14	6
Insula	R	43	9.13	46	9	2
Escapability parametric modulation - : Fast Predator						
Parahippocampal Gyrus	L	24	4.98	-21	-12	-24
Parahippocampal Gyrus	R	19	4.21	39	-39	-24
Middle Frontal Gyrus	L	48	7.06	-9	60	6
Escapability parametric modulation + : Slow Predator						
Middle Temporal Gyrus	L	86	5.62	-57	-35	-6
Insula	R	44	6.23	39	-14	8
Superior Occipital Gyrus	L	30	4.57	-14	-89	19
Superior Occipital Gyrus	R	54	4.24	17	-88	19
Escapability parametric modulation - : Slow Predator						
Middle Frontal Gyrus	L	84	7.28	-30	60	15
ACC	R	35	8.29	18	42	0
Caudate	L	14	4.34	-9	-51	27
Caudate	R	20	4.92	14	15	-3

Table S5. Activation Table for Contrast [High Reward > Low Reward]

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates	
				x	y
High Reward > Low Reward					
Putamen	L	57	8.42	-28	-3
Putamen	R	50	9.92	36	-12
Middle temporal gyrus	L	62	11.07	-60	-6
Middle temporal gyrus	R	56	9.28	54	-9
Inferior frontal gyrus	L	49	7.38	-51	30
Superior frontal gyrus	R	66	7.35	3	54

Table S6. Activation Table for Contrast [High Shock > Low Shock]

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
High Shock > Low Shock						
Superior temporal gyrus	L	147	11.16	-58	3	6
Insula	L	54	7.36	-41	-14	3
Insula	R	35	6.04	39	3	-12
MCC	R	32	5.96	6	-9	39
Parahippocampal gyrus	L	10	9.78	-24	-18	-9
Hippocampus	R	20	7.55	36	-15	-18

Table S7. Activation Table for Contrast [High Shock > Low Shock] (Fast Predator)

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
High Shock > Low Shock (fast predator)						
Midbrain	L	190	4.95	-1	-32	-13
ACC	R	179	4.36	3	25	7
Precuneus	L	335	4.03	-27	-54	8

Table S8. Activation Table for Contrast [High Shock > Low Shock] (Slow Predator)

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
High Shock > Low Shock (slow predator)						
Medial temporal gyrus	L	161	4.64	-42	-33	0
Insula	R	25	4.04	40	-16	10
Hippocampus	R	36	4.50	21	-30	-12
Superior temporal gyrus	R	26	4.30	45	-12	-3
Amygdala	R	17	4.11	24	4	-20

Table S9. Activation Table for Parametric Modulation with Reward Preference

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Parametric modulation with reward preference (fast predator)						
Fusiform gyrus	L	35	5.75	-42	-6	-27
Inferior temporal gyrus	R	23	4.93	60	-9	-18
Insula	L	80	5.04	-39	-12	6
Putamen	L	42	4.86	-28	-18	3
Middle temporal gyrus	L	81	4.11	-60	-57	-3
Putamen	R	78	4.88	33	-6	6
Parametric modulation with reward preference (slow predator)						
Middle occipital gyrus	L	85	5.86	-36	-66	0
Caudate	R	50	6.70	15	-3	15
Calcarine	R	38	4.79	18	-48	3
Insula	R	35	4.48	33	24	3
Caudate	L	46	4.88	-9	3	9

Table S10. Activation Table for Parametric Modulation with Shock Avoidance

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Parametric modulation with shock avoidance (fast predator)						
Inferior temporal gyrus	R	21	4.40	51	-60	-18
Midbrain	R	10	4.93	21	-18	-15
Superior temporal gyrus	R	21	4.83	60	-33	12
Inferior frontal gyrus	L	14	4.16	-45	42	15
Parametric modulation with shock avoidance (slow predator)						
PCC	L	23	4.92	-12	-39	6
Thalamus	R	35	6.27	0	-18	15
Middle temporal gyrus	L	45	4.69	-42	-66	15
Superior temporal gyrus	L	48	5.89	-60	-45	18

Table S11. Activation Table for Parametric Modulation with Bayesian optimality

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
Parametric modulation with Bayesian optimality (fast predator)						
MCC	R	24	4.31	6	33	21
Superior frontal gyrus	L	63	4.89	-24	9	54
Superior motor area	R	18	4.33	12	9	54
Parametric modulation with Bayesian optimality (slow predator)						
Hippocampus	R	20	4.18	33	-33	-3
Middle occipital gyrus	R	12	4.40	36	-87	-3
Precentral gyrus	R	41	4.32	63	-3	24
Precentral gyrus	L	18	4.05	-51	-3	24

Table S12. Activation Table for Connectivity Analysis

P<0.05, FDR corrected

Brain Region	Left/Right	Cluster Size	t-score	Coordinates		
				x	y	z
MCC Seed						
Midbrain	L	19	4.20	-5	-30	-13
Thalamus	L	25	3.97	-15	-13	6
Thalamus	R	30	4.15	18	-19	1
Hippocampus Seed						
PCC	R	20	4.19	4	-47	26

SI Text

Acquisition and Analysis of fMRI data

All fMRI data were acquired using a GE Discovery MR750 3.0 T scanner with 32-channel headcoil. The imaging session consisted of two function scans, each twenty minutes, as well as a high-resolution anatomical T1-weighted image (1mm isotropic resolution) collected at the beginning of each scan session. For functional imaging, interleaved T2*-weighted gradient-echo echo planar imaging (EPI) sequences were used to produce 45 3-mm-thick oblique axial slices (TR = 2 sec., TE = 25 ms, flip angle = 77°, FOV = 192 x 192 mm, matrix = 64 x 64). Each functional run began with five volumes (1000 msec) before the first stimulus onset. These volumes were discarded before entering analysis to allow for magnetic field equilibration. Stimulus were presented using Cogent (matlab-based package). Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement.

Analysis of fMRI data was carried out using scripted batches in SPM8 software (Welcome Trust Centre for Neuroimaging, London, UK; http://www/fil.ion.ucl.ac.uk/spm) implemented in Matlab 7 (The MathWorks Inc., Natick MA). Structural images were subjected to the unified segmentation algorithm implemented in SPM8, yielding discrete cosine transform spatial warping coefficients used to normalize each individual's data into MNI space. Functional data were first corrected for slice timing difference, and subsequently realigned to account for head movements. Normalized data were finally smoothed with a 6-mm FWHM Gaussian kernel.

Preprocessed imaged were subjected to a two-level general linear model using SPM8. The first level contained the following regressors of interest, each convolved with the canonical two-gamma hemodynamic response function: a 2-second box-car function for the onset of the trial (where the color of the incoming predator is shown); a 4-8 second (duration jittered) box-car function from the onset to 2s before when subjects make the flight decision; a 2-second boxcar (function for the phase before subjects make the flight decision; a 4-8 second (duration jittered) box-car function for the remainder of the trial. Mean-centered trait anxiety ratings, escapability

ratings and parameters in the Bayesian decision model were included as orthogonal regressors. In addition, regressors of no interest consisted of motion parameters determined during preprocessing, their first temporal derivative and discrete cosine transform-based temporal low frequency drift regressors with a cutoff of 192-seconds.

Beta maps were used to create linear contrast maps, which were then subjected to second-level, random-effects one-sample t tests. In Addition, A flexible factorial model was used to examine the main effects of predator type, reward level and shock level. Interaction effects between predator type, reward level and shock level were also examined using the factorial model. The resulting statistical maps were thresholded at $P < 0.05$ corrected for multiple comparisons (false discovery rate [FDR] corrected (41)). A flexible factorial model was used to examine the interaction effects between predator type, reward level and shock level. The threshold for those specific contrasts was set at $p < 0.05$ (FDR corrected).

A hypothesis driven regions of interest (ROI) analysis was performed after the whole brain analysis for regions with strong a priori spatial hypotheses. The ROI analysis was performed using regions associated with the processing of fear, threat and decision making. Independent ROIs were chosen from previous research showing similar effects (20, 42). The threshold for these analyses was set at $p < 0.05$, small volume correction (SVC).

The functional connectivity analysis was performed for the response phase (escape decision) using a generalized psychophysiological interactions (PPI) approach. The connectivity analysis was carried out based on the [predator condition > control condition] contrast.

Bayesian Decision Making Model

The empirical rationale of the model set up can be found in SI Appendix. A priori to observing attacking, the attack distance of a certain predator is believed to be drawn from a Gaussian distribution $AD|c \sim N\left(\mu^{(c)}, \sigma^2\right)$, where c represents the predator type. At the start of the experiment, the mean parameters are unknown and hence assumed to follow the same prior

distribution. Here we adopt the conjugate prior distribution $\mu^{(c)} \sim N(\mu_0, \sigma_0^2)$ with a large variance to reflect minimum prior knowledge. Meanwhile, we assume the variance of likelihood (σ^2) to be known, because in the practice phase subjects have already been exposed to predators with identical AD variance as in the formal experiment.

Upon observing attacks, the posterior distribution for the mean parameter is updated by the Bayes rule, yielding:

$$p\left(\mu^{(c)} | \left\{AD_i^{(c)}\right\}\right) \propto N\left(\mu^{(c)} \middle| \mu_0, \sigma_0^2\right) \prod_i N\left(AD_i^{(c)} \middle| \mu^{(c)}, \sigma_0^2\right) = N\left(\mu^{(c)} \middle| \mu_n^{(c)}, \sigma_n^{(c)2}\right),$$

where $\left\{AD_i^{(c)}\right\}$ are the *observed* distances of a total number of n attacks from type-c predators. The posterior is also a Gaussian, with parameters updated through $1/\sigma_n^{(c)2} = 1/\sigma_0^2 + n^{(c)}/\sigma^2$ and $\mu_n^{(c)} = \sigma_n^{(c)2}\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_i AD_i^{(c)}}{\sigma^2}\right)$. This posterior of the mean parameter directly *induces* the (posterior) predictive distribution of the upcoming attack distance, given by

$$p\left(AD_{n+1}^{(c)} \middle| \left\{AD_i^{(c)}\right\}\right) = \int N\left(AD_{n+1}^{(c)} \middle| \mu^{(c)}, \sigma_0^2\right) p\left(\mu^{(c)} | \left\{AD_i^{(c)}\right\}\right) d\mu^{(c)} = N\left(\mu_n^{(c)}, \sigma_n^{(c)2} + \sigma^2\right).$$

An ideal Bayesian learner will base its FID choice on this distribution. Now, under the context of the current paradigm, a subject chooses FID from a finite set of options by trading off two critical factors: the *risk of getting shocked* and the *monetary reward*. With a large FID, risk is reduced while less reward will be given; with a small FID, the opposite. We then define an overall utility as a weighted combination of the two factors:

$$u(FID, AD) = \beta_1 \boldsymbol{I}(caught) + \beta_2 M(FID)\big(1 - \boldsymbol{I}(caught)\big).$$

The coefficients β_1, β_2 are individuals-specific weights to adjust the preference between the two factors. $\boldsymbol{I}(caught)$ is the indicator function for the event of getting caught and killed (evaluates to 1 if caught, 0 otherwise); and $M(FID)$ is the amount of money rewarded if escape is successful. This utility is a random function since getting caught is a random event --- the optimal decision should then be based on the *expected value* of utility, namely the *Bayesian risk*, which is estimated from the latest posterior predictive distribution and takes the form of

$$U(FID) = \mathbb{E}\, u(FID, AD) = \beta_1 \Pr(caught|FID) + \beta_2 M(FID)\big(1 - \Pr(caught|FID)\big).$$

Here the probability of being caught $\Pr(caught|FID)$ can be computed by solving a simple chasing problem, where the actual speed of the particular predator and subject were taken into computation to determine if the subject would be caught or not.

The Bayesian learner's optimal choice is set as a reference to measure the performance of subjects. Clearly, the optimal FID is one that maximizes $U(FID)$. Yet, the behavior of a human player is also influenced by unobserved factors such as personality traits and is not necessarily Bayes optimal. To quantify individual differences of decision making through coefficients (β_1, β_2), we fit a discrete choice model (multinomial logit)(43) for each subject, which assumes that the probability of picking an option is proportional to the corresponding exponential utility:

$$\Pr(FID = x) = \frac{\exp\big(U(FID = x)\big)}{\sum_{y \in choices} \exp\big(U(FID = y)\big)}.$$

The coefficients can be estimated by maximizing the overall likelihood, namely

$$(\hat{\beta}_1, \hat{\beta}_2) = argmax \prod_i \Pr(FID_i | previous\ AD's),$$

where the product is taken over trials. By fitting to our data, a subject can be quantified by the corresponding coefficients:

Pain avoidance: expected to be non-positive, and a bigger absolute value means stronger aversion towards risk and its associated penalty (electric shock).

Reward Preference: expected to be non-negative, and a bigger value measures stronger favor of monetary reward.

Those parameters are then entered into fMRI parametric modulation analysis to determine the brain regions where signals covariate with the parameters.

On every particular trial, the difference between the actual utility $U(FID)$ and the Bayesian optimal utility $U(FID)_{max}$ is calculated as a measure of choice optimality. The "utility" used to calculated differences here can be "normalized" by dividing out β_2, the reward preference parameter.

To account for the effect from varying shock calibration levels, we tested correlation between the shock/reward beta values and the calibration level. Individual subjects' shock calibration level does not correlate with either shock avoidance beta (r = .19, p = .36) or reward preference beta (r = .11, p = .58)