How is Emotional Evidence from Multiple Sources Used in Perceptual Decision Making?

Hilary H.T. Ngai[1], Janet H. Hsiao [1,2,3], Christian Luhmann[4], Aprajita Mohanty[4], Jingwen Jin[1,2]

[1]Department of Psychology, The University of Hong Kong, The University of Hong Kong, Hong Kong SAR, China
[2]State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR, China
[3]Institute of Data Science, The University of Hong Kong, Hong Kong SAR, China
[4]Department of Psychology, Stony Brook University, New York, USA

Running title: Multi-Evidence Emotion Perceptual Decision Making

Author Note

This work has been presented at The Society for Affective Science Annual Conference 2021 & 2022. This work is funded by the University of Hong Kong, Seed Fund for Basic Research (104005616.106012.30200.301.01 awarded to JJ).

Corresponding: **Jingwen Jin, Ph.D.,** mailto:Frances.Jin@HKU.HK The University of Hong Kong

Acknowledgement: The authors thank Dr. Antoni B Chan and Anita Liao for their help in eye movement data management and thank the research assistants who have helped data collection.

Conflicts of interest: The authors declare no conflict of interest.

Preregistration and open science
We pre-registered the experiment on the Open Science Framework (OSF). Pre-registrations, materials are made available on the OSF at https://osf.io/e8jvb and https://osf.io/9dpcj. Data can be made available upon request to the corresponding author.

Abstract

Judging the overall emotional nature of a scene often requires us to deliberately integrate pieces of evidence that range from neutral to emotional in nature. Research so far has focused on relatively rapid decisions. Hence, how evidence is valuated and sampled given ample time, and the precise computations involved in integrating emotion-related evidence to form an overall perceptual decision remain unclear. Further, it is unclear how this process is influenced by decision-makers' endogenous (voluntary) attention directed towards emotional evidence. In two experiments, human subjects (N=33 and N = 40) viewed arrays of faces with expressions ranging from extremely fearful to extremely happy. They decided whether the faces were more fearful or happier on average in the absence (Experiment 1) and presence (Experiment 2) of endogenous attention directed towards the fearful or happy faces. Results showed that subjects evaluated the same emotional faces differently depending on the face array context. While they attentionally sampled extreme emotional evidence more, they did not weigh it higher or lower in their decision making. Rather, their decision-making strategy was best modeled by weighing all pieces of context-dependent subjective evaluated emotional evidence equally. Furthermore, endogenous attention influenced sampling of the emotional evidence, but the decision-making strategy remained unchanged. In conclusion, the findings show that when forming a judgment about multiple emotional stimuli without time pressure, humans weigh multiple pieces of subjectively valuated context-dependent evidence equally. These findings elucidate specific computational mechanisms of emotion-related decision making, shedding light on the mechanisms of normal and pathological affective decisions.

Key words: emotion, multi-element, face array, perceptual decision-making, computational modelling

Introduction

In daily life, emotional stimuli are rarely encountered in isolation or unpredictably. Rather, we encounter these stimuli in busy and noisy contexts consisting of a plethora of other stimuli. We also tend to experience them in contexts regarding which we have prior knowledge about what stimuli are relevant or likely (Mohanty et al., 2023). Emotion-related perceptual decision making thus requires the observer to deliberatively sample and integrate multiple pieces of evidence to form an overall judgement, a process that can also be heavily influenced by top-down factors such as attention and goals of the decision maker (Bacon & Egeth, 1994; Hahn & Gronlund, 2007; Sussman, Jin, et al., 2016). Such deliberate decision making is critical for adaptive interactions with our environments because errors in this process can yield biased impressions with detrimental personal and social consequences. So far, the precise computation of how humans valuate, sample, and weigh multiple pieces of emotional evidence in deliberate perceptual decision making remains unclear. Do they treat all evidence equally? Or discount extreme evidence as outliers? Or weigh extreme evidence higher due to its emotional saliency? How is this process influenced by attention when only some of the evidence is goal-relevant? The present study uses a novel behavioural paradigm, coupled with eye tracking and computational modelling to understand the process by which multiple pieces of emotional evidence is deliberately integrated for decision making.

Perhaps due to their relevance for survival, studies have typically focused on split second perceptual decisions regarding emotional stimuli (Elias et al., 2017; Goldenberg et al., 2020; Haberman & Whitney, 2009, 2010; Ji et al., 2014; Li et al., 2016; Sun & Chong, 2020; Wolfe et al., 2015). However, many natural perceptual decisions unfold over much longer durations (e.g., multiple seconds). Given ample time, deliberate perceptual decision making can be drastically different, as has been proven in economic decision making (Spektor et al., 2021). When making

rapid perceptual decisions about multiple pieces of evidence, the ensemble coding model shows that humans efficiently extract the "gist", or summary statistics, across all pieces of evidence available in a complex scene (Alvarez, 2011; Ariely, 2001; Cavanagh, 2001; Chong & Treisman, 2003; Dakin & Watt, 1997; Leib et al., 2014; Parkes et al., 2001; Watamaniuk & Duchon, 1992; Watamaniuk et al., 1989; Whitney & Yamanashi Leib, 2018) using distributed attention (Alvarez & Oliva, 2008; Elias et al., 2017; Haberman & Whitney, 2009, 2010; Hubert-Wallander & Boynton, 2015; Li et al., 2016; Sun & Chong, 2020; Wolfe et al., 2015). When integrating evidence, they engage in "robust averaging" (de Gardelle & Summerfield, 2011; Huber, 2004; Li et al., 2017; Vandormael et al., 2017), downweighing evidence with extreme values because they are considered outliers or unreliable (i.e., farther from the expectation) (Alvarez, 2011; Epstein et al., 2020; Whitney & Yamanashi Leib, 2018). Notably, such decision performance is also affected by the variance of evidence displayed, because variance signifies the overall reliability of the estimated average of a set of evidence (de Gardelle & Summerfield, 2011; Haberman et al., 2015; Michael et al., 2014; Whitney & Yamanashi Leib, 2018). A similar downweighing of extreme evidence has been observed when rapidly extracting the average of multiple emotional stimuli (Haberman et al., 2009; Haberman & Whitney, 2009).

In contrast to rapid decision making, given ample time, deliberate perceptual decision making may involve different computational processes, particularly due to the saliency of emotional stimuli. First, with the time constraint lifted, people can register each piece of evidence properly, and carry out *purposeful* and *active* sampling and *resampling* of the evidence space by shifting overt attention (Gidlöf et al., 2013; Itti & Koch, 2000; Kietzmann et al., 2011; Posner, 1980; Rahn et al., 2016; Wyart & Tallon-Baudry, 2008). Second, more extreme (or intense) emotional stimuli are more likely to be considered salient (Feldman, 1995; Posner et al., 2005) and

receive more attention (Calvo & Nummenmaa, 2016; Kulke, 2019; Mogg & Bradley, 1999; Nummenmaa et al., 2006). Importantly, stimuli that attract more attention are more likely to be chosen (Armel et al., 2008; Kaanders et al., 2022; Krajbich & Rangel, 2011; Sepulveda et al., 2020). Moreover, it has been proposed that stimuli which attracted more visual attention are weighed higher in the decision (Krajbich et al., 2010; Orquin et al., 2021; Orquin & Loose, 2013; Pärnamets et al., 2016; Thomas et al., 2019). Together, it is thus plausible that when making deliberate emotion-related perceptual decisions, extreme evidence, instead of being neglected or downweighed, may play an equal or even larger role in forming the decision.

To understand how emotion-related decisions are made, it is important to examine every stage of a multi-stage perceptual decision-making process **(Figure 1)**. First, the decision-maker converts each piece of external information into an internal representation (Gold & Shadlen, 2007), which may vary across individuals and within an individual across contexts. Indeed, judgements regarding the degree of extremity of the emotional evidence can be quite subjective and dependent on the local or global context (Barrett, 2006; Nummenmaa et al., 2018; Russell, 2003). It has been shown that subjective valuations of emotional stimuli are sensitive to relative differences between stimuli (Jin et al., 2015), and may be placed onto different scales based on contexts, e.g., when all pieces of evidence are quite similar versus highly variable (Clark & Lawless, 1994; Lawless & Clark, 1992). Hence, for an in-depth analysis of integration of emotional evidence it is imperative to examine the contribution of the observer's contextually based evaluation of emotional intensity (i.e., the subjective value) rather than just objective values (e.g., based on physical properties or population ratings). Second, the decision-makers sample the evidence space regarding their choice alternatives. Due to their salience, emotional stimuli capture attention in a "bottom-up" stimulus-driven manner and tend to be sampled more (Feldman, 1995; Orquin & Loose, 2013). Considerable

research from the field of behavior economics (Hertwig & Erev, 2009) and basic perception (Kaanders et al., 2022; Orquin et al., 2021; Orquin & Loose, 2013) shows that how evidence is attentionally sampled may contribute to its under or overweighing in decision making, highlighting the importance of examining this step in the integration of emotional evidence. Finally, observers compute a decision variable as a weighted sum of evidence values and compare it to a decision criterion, executing a response if the variable crosses the decision criterion. Currently, it is unknown how emotional evidence is weighed in this computation. Specifically, it is unclear whether this weighing involves the objective values, subjective values and/or the attention sampling of the emotional evidence.

While the role of emotional evidence described above focuses on bottom-up capture of attention, recent advances in affective science has shown that endogenous attention or expectation modulates sensory processing in a top-down manner (Barrett & Simmons, 2015; Bhatia, 2014; Feldman Barrett, 2011; Mohanty & Sussman, 2013; Sussman, Jin, et al., 2016; Sussman, Szekely, et al., 2016; Sussman et al., 2017). Endogenous emotion-related attention (e.g., attending to fearful or happy stimuli) dictates what emotional evidence is relevant versus irrelevant, affecting subsequent perceptual decision making. For example, the same set of fearful and neutral stimuli are better discriminated when attention is directed to fearful compared to when attention is directed to neutral stimuli (Glasgow et al., 2022; Sussman et al., 2017). In a multi-evidence situation, endogenous attention towards a particular emotion may further increase the saliency of the task-relevant extreme evidence. Manipulating endogenous attention may hence alter the decision strategy when sampling and integrating provided evidence, leading to biased decisions. However, such potential effects of top-down attention modulation have yet to be examined in deliberate multi-evidence emotion-related decision making.

The overarching aim of this study is to (a) gain a mechanistic insight into how multiple pieces of emotional evidence are valuated, sampled and integrated when deliberatively forming a perceptual decision, and (b) investigate how endogenous emotion-related attention further affects this decision-making process. We conducted two experiments in which we tested whether given enough time, extreme emotional evidence would be underweighed in decision making, as is the case of rapid decision making. Unlike most earlier studies, we tested our predictions by carefully measuring contextually influenced subjective evaluation of emotional evidence as well as attentional sampling of this evidence. We demonstrated that humans evaluated the same emotional evidence differently and adopted an internal scale that was sensitive to the context surrounding the evidence. While they sampled extreme emotional evidence more, they did not weigh it higher or lower in their decision making. Their decision-making strategy was best modeled by weighing all pieces of evidence equally based on their context-dependent subjective valuations of the emotional evidence. Finally, we demonstrated that while endogenous attention influences sampling of the emotional evidence, the decision-making strategy remained unchanged, continuing to rely on equal weighing of subjectively evaluated evidence. Overall, we show that when making an overall decision regarding multiple stimuli in the absence of time pressure, humans are able to weigh multiple pieces of subjectively valuated context-dependent evidence equally, allowing them to arrive at a non-biased and rational decision given their context.

Experiment 1

In Experiment 1 we used a multi-stage approach to gain insight into how emotional evidence contributes to each stage of perceptual decision making, starting with the initial evaluation of the evidence, and ending with the integration of the evidence in forming the decision variable. We recorded eye movement data as subjects viewed multiple emotional faces ranging

from extremely fearful to extremely happy before deciding whether the array of faces was more fearful or happier on average. The emotion intensity is coded as the objective emotion values (OEV, detailed later). Across trials, the face arrays differed in variance such that their OEVs vary from clustering around midpoint between fearful and happy (low variance) to spanning a wide range (high variance). We also recorded participants' subjective valuations of fearfulness/happiness of each emotional face within the face array context. This was the subjective emotion values (SEV).

We hypothesized that the SEV of each emotional face would be affected by the trial context, such that subjects may perceive the difference in emotional intensity as more evident, manifesting as a wider range of SEV relative to the OEV in low compared to high variance trials. For evidence sampling, we used eye movement to examine how the emotional evidence in the face array is sampled as a function of its SEV, because eye movement is a suitable proxy for overt visual attention and is sensitive to the goal-relevance and value of targets (Grebitus et al., 2015; Krajbich et al., 2010; Krajbich & Rangel, 2011; Kulke, 2019; Sepulveda et al., 2020). We hypothesized that emotional faces with more extreme SEV would attract more visual attention, which would manifest as a U-shape pattern of eye movements as a function of faces ordered from extremely fearful to extremely happy. This effect would be heightened for high versus low variance condition. Finally, we hypothesized that the decision maker will integrate the sampled subjectively-valuated evidence to form a Subjective Decision Value (SDV)[1] (Gold & Shadlen, 2007), such that the strength of the SDV will be converted into the probability of choosing one of the options, yielding a binary decision (with some level of uncertainty). Critically, the formation of the SDV was hypothesized to be influenced by individuals' subjective valuation and attention rather than

[1] The decision variable (often referred to as DV) taking on different values differs in different contexts, e.g., neuronal firing rate or amount of money. We use subjective decision value (SDV) to highlight that this is an internally computed value that is intrinsically subjective.

faithfully representing the OEV. We tested this last hypothesis by examining models in which the SDV was computed based simply on average OEV (model 1), average SEV (model 2) or SEV weighted by different eye movement measures (models 3-5). Embedded in the latter 3 models was the idea that the extreme emotional evidence may be further upweighed if it was attentionally sampled more via eye movements. Comparing models 2-5 with model 1 would reveal whether subjectivity in the valuation of evidence plays a critical role in decisions. Comparing models 3-5 with model 2 would reveal whether attention allocation was a mere reflection of SEV (if model 2 was the best fitting model) or attention had additional contribution to the decision by influencing how evidence was weighed (if any of models 3-5 was the best fitting model). We also hypothesized that lower compared to higher variance condition would lead to worse model-fitting performance, as the pieces of emotional evidence were less differentiable (more homogenous) and hence there would be a stronger effect of stochasticity on the binary decisions.

Subjects

Forty-six subjects (*Mean Age* = 20.4 years old, *SD* = 3.0 years old; Women = 71.7%) from the University of Hong Kong were recruited. The current sample size was determined based on previous research (de Gardelle & Summerfield, 2011). In this study, their sample sizes for four different variations of a similar task were 30, 14, 16 and 24 respectively. Data from thirteen subjects were discarded: four due to incomplete experiment, five due to technical failure and four due to outliers in eye movement data (see Eye Data Processing). Thus, the final dataset consisted of thirty-three subjects. All subjects provided written informed consent, reported fluency in English, and had normal or corrected-to-normal vision. The study was approved by the Human Research Ethics Committee of the University of Hong Kong. Subjects were either given partial course credit or were compensated 50HKD per hour.

Stimuli

Image stimuli consisted of 10 (five fearful and five happy) faces selected from five Asian actors (two female) from the HKU Face Database (Zhang et al., 2019). Asian face models were chosen due to the demographics of the subjects and to increase sensitivity to the local sociocultural context (Masuda et al., 2008). The images were first converted from color to grayscale and cropped around the chin and ears (300x300 pixels). Fearful and happy faces were chosen because these two emotions are high arousal yet of opposite valence (Watson et al., 1999). Although both anger and fear are of negative valence, angry faces were not chosen, because anger provokes behavioral changes in the target of that anger (e.g., the subjects), eliciting a more complex response (Pichon et al., 2009).

To create an array of faces ranging from fearful to happy, we used a face morphing software FantaMorph (Abrosoft Fantamorph Version 5.6.2; www.fantamorph.com). For each face model, facial features such as the corner of eyebrows, shape of the eye, nose and mouth were matched for their fearful and happy expressions. Then, 99 synthetic images were created by linearly interpolating between the fearful and happy facial images, yielding 101 images for each face model, where each image depicted varying degrees of fearful/happiness. We numerically coded the 101 images, with an emotional value unit of 0 corresponding to the face model's original most fearful face and 100 corresponding to the happiest face. These emotional value units are hereby referred to as the OEV of each face image for the remainder of this paper. Note, that there is a lack of universal emotion expression value scale, as there would be if using non-emotional stimuli such as color (i.e., RGB values).

Next, because fearful and happy facial expressions tend to be open-mouthed, ghosting effects may occur when morphed due to the presence of teeth. Adobe Photoshop (Version 22.1.1)

was thus used to manually edit these images to create natural-looking faces. Lastly, the 505 (101 fearful-to-happy images x 5 face models) stimuli were equalized for contrast and luminance using the SHINE Toolbox on MATLAB (Willenbockel et al., 2010). We selected the default functions (specMatch and histMatch) to equate the Fourier spectra followed by the histograms matched to the whole image with one iteration.

Experimental Procedure

All subjects were asked to complete two tasks. In the first perceptual decision-making task, subjects' eye movements were tracked during the 8000-msec viewing of the multi-evidence emotional stimuli of each trial (**Figure 2A**). The second task was a subjective emotion rating task with no time constraints (**Figure 2C**). Both tasks comprised of 75 trials each and each trial presented 9 faces.

These 75 trials were divided into 3 conditions, each of which reflected a different variance level (high, medium, low variance), with 25 trials in each condition. The 75 trials were pseudorandomized such that the same face model and variance level were not presented more than four trials in a row, to avoid local learning effects. In each trial, nine images from the same model were presented. Eight of these images formed the multi-evidence set, with varying OEV. In the center of the multi-evidence set was the ninth image — the image with the OEV of 50 for that face model. Faces in the three different variance levels were generated by drawing eight OEVs from a Gaussian distribution, with a mean of 50 (i.e., the midpoint of the fearful-to-happy spectrum of OEV), and a standard deviation of either 10 (low variance), 20 (medium variance), or 30 (high variance). For example, in a low variance trial, the eight faces would likely have OEVs close to 50. In contrast, a high variance trial would likely to have faces covering a wide range of the OEV

values. After the OEVs of the eight faces were determined for each trial, the location of the 8 faces were randomly assigned into eight placeholders in an annulus format.

The decision-making task. In the decision-making task **(Figure 2A)**, each trial began with a drift correction check where a solid circle appeared in the center of the screen, and the stimuli were presented after the subjects' fixation was within one degree of visual angle. Nine empty squares then appeared for 1000 msec, which served as placeholders for subsequently presented face images. Then, the set of nine faces (the 8 faces in the annulus and the 9th "reference" face in the center) replaced their respective placeholders for 8000 msec, during which subjects freely viewed the images. Subjects were instructed that the central "reference face" served as a benchmark for the face model when they were 50% fearful and 50% happy. They were encouraged to use the reference face as a criterion for their upcoming decision but instructed *not* to include the reference face in their decision. The purpose of this reference face was solely to provide a benchmark for subjects to apply the decision rule by setting a common criterion when making their binary decision. Note that our analyses were all within-subject, and subjective decision criterion was estimated at the subject- or variance-level (see statistical analyses). Thus, how reliably subjects used this reference face did not influence the results.

After the 8000 msec of viewing, subjects were instructed to indicate whether the eight faces were, on average, more fearful or happier, and to do so as quickly as possible. We chose a relatively long stimuli presentation duration (8000 msec) for ecological validity, as it is more likely that individual sample and integrate evidence over longer time durations in daily life (Buck, 1980, 1988). This duration also facilitated the estimation of eye movement trajectory described below. The decision-making task was delivered using Experiment Builder (Version 2.2.61) compatible with eye-tracker software.

The subjective emotion rating task. In the subjective emotion rating task **(Figure 2C)**, subjects were presented with the same image arrays in the same trial sequence used in the decision-making task but were instructed to rate each of the eight faces shown on a scale ranging from 0 ("very fearful") to 10 ("very happy"). Once again, the subjects were instructed that the central reference face served simply as a benchmark of that face model's 50% fearful and 50% happy face and they did not need to rate it. Ratings obtained from this task will be referred to as the SEV. The inclusion of this second task adds value to the experiment by effectively isolating the role of SEV from the role of evidence weighing and sampling. This is different from previous studies in which the value of each element is defined either by the objective value provided by the experimenter, or a subjective rating obtained outside of the trial context. This two-task design allowed us to examine each stage in the decision-making process to help elucidate its contribution to the overall decision-making process. The subjective emotion rating task was delivered using PsychoPy (Version 3.2.4).

Apparatus

Eye movement of each subject's dominant eye was recorded by an EyeLink 1000 eye-tracker (SR Research Ltd.). Subjects sat 60 cm in front of a 22″ monitor (1024 × 768 pixels). Using pupil and corneal reflection, the system has a sampling rate of 1000 Hz which was suitable for monocular sampling. Eyelink default settings (i.e. saccade motion threshold of 0.1° visual angle, saccade acceleration threshold of 8000°/s^2 and saccade velocity threshold of 30°) were used for data collection. At the beginning of the eye-tracking task, default nine-point calibration and validation were conducted and repeated until the drift correction error was less than one degree of visual angle. All subjects used a chin rest to reduce head movements.

Eye Movement Data Processing

When approaching the eye movement data, we considered both the spatial (i.e., fixation location) and temporal dimensions (i.e., duration, and sequence of transition amongst the fixation locations). Traditional eye movement analysis methods concerning the fixation count and fixation duration were calculated for each region of interest (ROI), which corresponded to the 9 placeholders within each trial. The 9 placeholders were defined by their corresponding coordinate vertices, creating a 180-by-180-pixel area to form a square for each. The ROIs for transition probability were optimized for the Eye Movement analysis with Hidden Markov Model (EMHMM) estimation.

Quality check was conducted for eye movement data by finding the mean number of fixations of all subjects and then detecting outliers by using the mean ± 2 SD as a boundary. Four subjects were discarded.

Eye Movement Analysis with Hidden Markov Models (EMHMM). To investigate both the spatial and temporal dimension, we employed a data-driven, machine-learning approach using the EMHMM MATLAB toolbox (version 0.77). EMHMM uses Hidden Markov Models (HMM) to analyze the observable time-series data provided from eye-tracking. EMHMM assumes that the ROI of each fixation is dependent only on the previously viewed ROI. The hidden states of an HMM correspond to the ROIs, and the observable fixation data is modeled as arising from an underlying dynamic process of the sequence of ROIs viewed (Chan et al., 2018; Chuk, Chan, et al., 2017; Chuk, Crookes, et al., 2017). Since our task includes a natural set of 9 ROIs of each trial, we predefined the spatial ROIs using the area of each face's location defined by the center coordinate and a radius of 90 pixels (i.e., 4 SD of a 2-dimensional Gaussian distribution) (Cho et al., 2022a; Cho et al., 2022b; Chuk et al., 2019). The prior probabilities across the 9 ROIs were

initialized as a flat distribution. Then, given the fixation point sequence from the eye movement data, the HMM estimates the probability of starting at a particular ROI (i.e. estimated prior probability) and the probability of fixations transitioning between a pair of ROIs (i.e. transition probability) (Chuk et al., 2014). All analyses regarding the eye movement data were conducted using MATLAB (Release 2020b; The MathWorks, Natick, MA).

Statistical Analyses

Examining valuation by comparing OEV and SEV. To investigate the relationship between each face's OEV and SEV, Spearman's rank correlations were computed. Correlations were calculated across all individual face images for each subject and each variance level separately. Using Fisher's z-transformation, the correlation coefficients were converted to *z* scores ($z = 0.5*\log((1+r)/(1-r))$) and a one-way analysis of variance (ANOVA) was applied at the group level to examine the effect of the variance level on the correlation between OEV and SEV.

We used the standard deviation (SD) to examine the spread of SEV versus OEV for each variance level. For SEV, we computed the SD of SEVs across the 8 faces per trial, per variance condition and per subject. The OEV (originally on a scale of 0-100) was first divided by 10 to match the scale of the SEV, and the SD of each trial's OEV was calculated. Then, for each trial, the difference between the spread of SEV and the spread of OEV was computed as SD of SEV minus SD of OEV, yielding a positive (negative) number if the spread of SEV is larger (smaller) than that of OEV. These difference scores were averaged across the 25 trials for each variance level and each subject. At the group level, three one-sample *t*-tests, one for each variance level, were applied to this mean difference to test whether the difference in SD was statistically significant.

Examining evidence sampling using eye movement data. To investigate how subjects sampled emotional evidence, the mean proportion of fixation counts, proportion of fixation durations, and transition probabilities to each ROI were analyzed. For each subject and for each trial, we grouped fixations by ROIs based on coordinate location, excluding fixations that fell outside the nine ROIs. Then, for each of the eight evidence-bearing ROIs (excluding the central reference face), the proportion of fixation counts, and duration values were computed. Finally, these fixation counts and durations for the eight ROIs were reordered based on the SEVs (in case of ROIs with equal SEVs within a trial, the ROIs were further reordered based on the OEV). After reordering, ROI1 always referred to the most fearful face of that trial, while ROI8 was the happiest face of that trial. Here onwards, we refer to this factor as the ROI Rank Order. In other words, the ROI Rank Order factor had 8 levels such that level 1 carried the eye movement information about the most fearful face for each trial, and level 8 carried the eye movement information about the happiest face for each trial. This re-ordering used the same rationale as in existing literature (de Gardelle & Summerfield, 2011).

The transition probabilities to each of the eight peripheral ROIs were computed by calculating the marginal transition probabilities. For each of the eight ROIs, this marginal transition probability equals the sum of all the transition probabilities to that ROI, weighed by the prior probability of the source ROI **(Equation 1)**. The prior and transition probabilities were the estimated outputs of the EMHMM.

$$p(ROI(j)) = \sum_{i=1}^{9} prior\ probability_i \times transition\ probability_{i \to j} \quad \text{(Equation 1)}$$

Note that the EMHMM was conducted using original ROI sequence (first central then clockwise) to preserve fidelity to the natural eye movement sequence, and only after computing

the marginal transition probabilities, were these probabilities reordered based on the ROI Rank Order outlined above.

For statistical analyses, trials were grouped according to variance levels. After taking the mean fixation count, fixation duration, and transition probabilities averaged across all trials within each variance level for each subject, repeated measures ANOVAs (rmANOVA) with 3 (variance level) by 8 (ROI rank order) factors were applied to compare the effects of variance level and ROI Rank Order on eye movement measures, specifically to examine if they followed the hypothesized quadratic or U-shaped pattern that varied with variance levels.

Computational modelling of the binary decision making. To gain deeper insight into how subjects integrated the emotion evidence, we compared five models differing only in how the SDV was computed (see **Supplementary Table 1**). In the first two models, all evidence was weighed equally and SDV was computed as the average OEV (Model 1) and average SEV (Model 2) presented in each trial. In the latter three models, SDV was calculated as a weighted sum of the SEV, based on the hypotheses that evidence integration may be further influenced by how eye movements (a proxy for visual attention) were allocated across the stimuli with various emotion values. Hence, SEV was weighed by one of the eye movement measures described above, excluding the proportion of eye movement towards the reference image: fixation count (Model3), fixation duration (Model 4), and marginal transition probability (Model 5) toward each ROI.

Each model assumed that the respective SDV generated probabilistic preferences by passing the SDV through a conventional softmax function **(Equation 2)**.

$$p(happy) = \frac{1}{1+ e^{-\mu(SDV-c)}} \quad \text{(Equation 2)}$$

We estimated the value of the SDV at which each subject was expected to be indifferent (i.e., p(happy)=0.5), *c*, and a determinism parameter that controlled how strongly related choices were to the SDV, μ. Parameter values were taken as those minimizing the negative log-likelihood of the choice data. These values were estimated separately for each subject and model using the optimization routines in the SciPy Python package (Virtanen et al., 2020). Following model fitting, the mean and median of each model's negative log-likelihood were calculated to summarize which model best described the subjects' task behavior (i.e., by which model yielded the lowest negative log-likelihood). Finally, to examine whether the overall pattern across all trials was consistently presented at each variance level, the trials were also grouped according to variance levels. Model fitting followed the same procedure. rmANOVA with 1 (variance level) factor was applied to compare the effect of variance level on the log likelihood of the best fitting model.

Experiment 1 Results

Effect of Evidence Variance on Subjective Emotion Valuation.

To examine whether SEV of each emotional face would be affected by the trial context, we examined the similarity between SEV and OEV as a function of variance level (**Figure 3A-C**). As expected, OEV and SEV were highly correlated with each other across subjects. Results from a one-way ANOVA demonstrated an effect of variance level on the Fisher's *z*-transformed correlation coefficient between SEV and OEV ($F(2, 64) = 270$, $p < .001$, $\eta_p^2=.894$). Tukey's post hoc criterion indicated that the average Fisher's correlation was lower in the low variance conditions (*Mean* = .788, *SD* = .0399) than both the medium variance condition (*Mean* = 1.130, $p < .001$) and high variance condition (*Mean* = 1.349, $p < .001$).

Moreover, it is clear from **Figure 3A** that when the variance was low, SEV spanned a larger range than OEV, indicating that the perceived dissimilarity in emotion values among the face

images was larger than the OEV. One-sample *t*-test confirmed that the difference in spread (measured as SD) between SEV and OEV was significantly greater than zero (M = .317, SD = .455) for trials of low variance, $t(32) = 4.00$, $p < .001$, $\eta_p^2 = .696$; close to zero (M = -0.0556, SD = 0.479) for trials of medium variance t(32) = -0.667, $p = .510$; and less than zero (M = -0.393, SD = 0.482) for trials of high variance $t(32) = -4.68$, $p < .001$, $\eta_p^2 = .815$ **(Figure 3D)**.

In sum, the behavioral results indicate that the perceived dissimilarity of emotional evidence is influenced by its context. Our results show that the perceived dissimilarity between faces was augmented compared to the OEV in low variance condition, whilst it was reduced compared to OEV in high variance condition. This pattern can also be interpreted as insensitivity of SEV to OEV variance, highlighting the importance of examining the contribution of subjects' internal representation of the external stimuli to the decision making.

Extreme Emotion were Sampled More.

Next, we examined how faces varying in SEV were sampled by examining the eye movement data (fixation count, fixation duration, and transition probability) towards each ROI (**Figure 4). Figure 4A** displays a sample subjects' eye movement trajectory in one trial, and a sample transition probability matrix generated by EMHMM is shown in **Figure 4B.** To test our hypothesis that the distribution of eye movement data (fixation count, duration and transition probability) **(Figures 4C-E)** will show an increasing quadratic effect of ROI Rank Order as variance increases, we examined the interaction of Rank Order by variance.

In addition to main effects of variance and ROI Rank Order (see **Supplementary materials**), there was a variance by ROI Rank Order interaction effect on fixation count $F(14,448) = 2.53$, $p = .002$, $\eta_p^2 = .073$, fixation duration $F(14,448) = 2.83$, $p < .001$, $\eta_p^2 = .081$, and transition probability $F(14,448) = 2.05$, $p = .013$, $\eta_p^2 = .060$, indicating that the effect of Rank Order varied

by variance. One-way ANOVAs conducted to probe the hypothesized U-shaped effects revealed significant quadratic effect in the low variance trials for fixation count $F_{quadratic}(1, 32) = 31.622$, $p < .001$, $\eta_p^2 = .497$, fixation duration $F_{quadratic}(1, 32) = 23.657$, $p < .001$, $\eta_p^2 = .425$ and transition probability $F_{quadratic}(1, 32) = 6.826$, $p = .014$, $\eta_p^2 = .176$, as well as a quartic effect for fixation count $F_{quartic}(1, 32) = 9.957$, $p = .003$, $\eta_p^2 = .237$, fixation duration $F_{quartic}(1, 32) = 11.206$, $p = .002$, $\eta_p^2 = .259$, and transition probability $F_{quartic}(1, 32) = 10.653$, p = .003, $\eta_p^2 = .250$. The quadratic effect of ROI Rank Order was also observed in medium variance for fixation count $F_{quadratic}(1, 32) = 12.657$, $p = .001$, $\eta_p^2 = .283$, fixation duration $F_{quadratic}(1, 32) = 13.827$, $p = .001$, $\eta_p^2 = .302$ and transition probability $F_{quadratic}(1, 32) = 6.549$, $p = .015$, $\eta_p^2 = .170$ while a quartic effect was observed in high variance for fixation count $F_{quartic}(1, 32) = 8.700$, $p = .006$, $\eta_p^2 = .214$, fixation duration $F_{quartic}(1, 32) = 15.483$, $p < .001$, $\eta_p^2 = .326$ and transition probability $F_{quartic}(1, 32) = 10.454$, $p = .003$, $\eta_p^2 = .246$.

Overall, in line with our hypothesis and contrary to findings using rapid decision-making tasks, present eye movement data demonstrated that faces with extreme SEV were sampled more rather than being ignored or filtered out. In addition, in the high variance condition, there was greater sampling of faces with extreme SEV as well as faces with the lowest SEV.

Evidence Integration depends on Subjective Valuation.

Finally, we examined how evidence is integrated to form a decision variable. We tested five models differing in how SDV was computed based on average OEV (Model 1) and average SEV (Model 2) as well as SEV weighted by eye movement fixation count (Model 3), fixation duration (Model 4), and transition probability (Model 5). Results indicated that two of the five models fit subjects' choice data better than the others **(Figure 5A)**. One was Model 2 in which the decision variable was the average SEV of each trial, and the other was Model 4, in which the SEV

of each image was weighted by its fixation duration proportion. The mean log likelihood implied a slight advantage for the former, and the median log likelihood implied a slight advantage for the latter. What was strikingly clear was that the choice model based on the OEV performed worse than any of the models involving SEV as seen from the high mean and median log likelihood in all trials and trials of different variance levels. Lastly, in the best fitting model based on the mean (Model 2), the trials of high variance (*Mean* = 12.6, *SD* = .0.421) had the lowest log-likelihood as compared to medium (*Mean* = 14.6, $p < .001$) and low variance trials (*Mean* = 13.7, $p = .189$) **(Figures 5B-D)** ($F(2, 64) = 7.67$, $p = .001$, $\eta_p^2 = .193$), indicating that the Model 2 best described the subjects' choice behavior in trials of high variance. This means that decisions in the low variance trials were more influenced by stochasticity.

Experiment 2

Experiment 1 established that SEV is different from OEV, higher SEV is attentionally sampled more as opposed to being downweighed or neglected, and that the perceptual decision is best captured by the simple averaging of SEVs of multiple emotional stimuli. In Experiment 2, we examined the influence of top-down factors, specifically endogenous emotion-related attention on the sampling and integration of multiple pieces of emotional evidence to arrive at a perceptual decision. Endogenous attention dictates what information is task- or goal-relevant, enhancing the saliency of the relevant information (Desimone & Duncan, 1995; Lang, 1997; Posner & Boies, 1971; Smith & Ratcliff, 2009; Taylor & Fragopanagos, 2005).

Experiment 2 included the same tasks as in Experiment 1; but in addition, participants' endogenous attention was directed towards a specific type of evidence (fearful or happy faces) to examine how it influences sampling and integration of this evidence. This endogenous-attention manipulation strategy has been used effectively in perceptual decision-making tasks with single

stimuli (Sussman, Szekely, et al., 2016; Sussman et al., 2017). We hypothesized that top-down attention towards a specific type of evidence (fearful and happy face-related attention) would lead to an imbalanced U-shape such that attention-congruent, compared to attention-incongruent emotional evidence, would attract more eye movement during sampling, more so for extreme evidence. Such an effect, measured by eye-movement, may further influence the subsequent formation of SDV, in the form of weights. As in Experiment 1, this hypothesis was tested by model comparison.

Subjects

Forty-six naïve young adults (Mean Age = 27.08 years old. SD = 9.79 years old; Women = 70%) from the University of Hong Kong were recruited in the experiment. Data from six subjects were discarded: two due to incomplete experiment, two due to technical failure and two due to outliers in eye movement data (same criteria as in Experiment 1). Thus, the final dataset consisted of forty subjects. Participation procedures, participation requirements and compensation were the same as Experiment 1.

Stimuli & Apparatus

Stimuli creation, selection and apparatus were the same as used in Experiment 1.

Experimental Procedure

The experiment procedure was the same as Experiment 1 with the following difference. In the decision-making task **(Figure 2B)**, endogenous attention was manipulated based on an existing paradigm, where the binary choice was in a “X or Not X” format (Sussman, Szekely, et al., 2016; Sussman et al., 2017), to alter the subjects’ attention across blocks (Imbriano et al., 2020; Summerfield et al., 2006; Sussman, Jin, et al., 2016).

The decision-making task. There were 90 trials in total, separated into 6 blocks, where the first 45 trials were the same as the latter 45. The main difference was that the instruction for binary choice was different across the repeated trials. Instead of deciding whether the faces were "more fearful or happy" on average as done in Experiment 1, in alternative blocks, subjects were asked to make a "Fearful or Not Fearful" decision under the (F)ear-attention condition or a "Happy or Not Happy" decision under the (H)appy-attention condition **(Figure 2B)**. This means the exact same display of evidence appeared once under F-attention and once under H-attention. The sequence of blocks was counterbalanced between subjects.

The subjective emotion-rating task. As in Experiment 1 **(Figure 2C)**, in Experiment 2, participants rated the emotions of each face in the 45 trials that were repeated twice in the decision-making task.

Data Processing and Statistical Analyses

Data was processed in the same way as Experiment 1. For analyses regarding eye movement and computational modelling, the trials under the same condition of endogenous attention manipulation were grouped together for the same analyses process as Experiment 1. For computational modelling analysis, "fearful" decisions included "fearful" decisions in the F-block and "not happy" decisions in the H-block. Whereas "happy" decisions included "happy" decisions in the H-block and "not fearful" decisions in the F-block.

Experiment 2 Results

Effect of Evidence Variance on Subjective Emotion Valuation was Replicated.

Experiment 2 replicated the SEV findings of Experiment 1, showing a clear distinction between subjects' internal representation (SEV) and the objective (OEV). The strength of OEV

and SEV association depended on the variance of the evidence. There was an effect of variance level on the Fisher's *z*-transformed correlation coefficient between SEV and OEV ($F(2, 78) = 296$, $p < .001$, $\eta_p^2=.884$). Tukey's post hoc criterion indicated that the average Fisher's correlation was lower in the low variance conditions (*Mean* = .871, *SD* = .0337) than both the medium variance condition (*Mean* = 1.285, $p < .001$) and high variance condition (*Mean* = 1.473, $p < .001$). Moreover, SEV was not sensitive to OEV variance, demonstrating an augmentation effect in low variance trials and a shrinkage effect in high variance trials **(Supplementary Figures 2A-C)**. One-sample t-test confirmed that the difference between SEV and OEV in terms of SD being greater than zero (*Mean* = .350, *SD* = .069) for trials of low variance, $t(39) = 5.07$, $p < .001$, $\eta_p^2 = .696$; close to zero (*Mean* = -0.0195, *SD* = .070) for trials of median variance t(39) = -0.278, $p = .782$; and less than zero (*Mean* = -.279, *SD* = .079) for trials of high variance $t(39) = -3.52$, $p = .001$, $\eta_p^2 = .808$ **(Supplementary Figure 2D)**.

Endogenous Attention Further Influences Attention Allocation.

Visual inspection of the eye-movement data showed that the U-shape results of Experiment 1 were replicated **(Figures 6A-C).** Given that all 3 levels of variance show a U-shape **(Supplementary Figure 3)**, we report the results pooling trials from all three variance level conditions. In addition, subjects sampled extreme emotions more, especially when congruent with the endogenous attention, showing a tilted U-shape in favor of evidence of the attended emotion. This was further confirmed by a significant interaction between endogenous attention and ROI in fixation count ($F(7,273)=4.89$, $p <.001$, $\eta_p^2=.111$), fixation duration ($F(7,273)=5.83$, $p<.001$, $\eta_p^2=.080$) and marginal transition probability ($F(7,273)=3.385$, $p =.002$, $\eta_p^2=.111$). **Figures 6A-C** show clearly that, across all measures, while the inlying ROIs (of less extreme values) show no

significant difference between F- and H-attention conditions, ROIs with more extreme values tend to be attended more when they were attention-congruent than incongruent.

Use of SEV in Evidence Integration Was Replicated.

Computational modelling results replicated that of Experiment 1 — Model 2, which had the average SEV of each trial as the decision variable, fit the subjects' choice data better than the other models as it had the lowest mean and median log likelihood **(Figure 6D)**. Similarly, Model 1 — based on the OEV, had the worst performance. Lastly, the trials of H-attention had lower log-likelihood in the best-fitting model (Model 2) as compared to F-attention trials ($F(1,39) = 8.57$, $p = .006$, $\eta_p^2 = .180$) **(Figures 6E-F)**, indicating that overall, the Model 2 better described the subjects' task behavior in trials under H-attention manipulation.

Discussion

In the real-world, we must frequently summarize information across several visual elements in our environment. For example, to judge whether the audience is enjoying a talk, the speaker may purposefully scan expressions of multiple faces and integrate such evidence over time to form a summary impression of whether the audience looks engaged or bored. Several aspects of how we make such decisions when we have ample time, as is often the case in real-life, remain unclear. How is the evidence subjectively evaluated differently depending on context and how is it sampled attentionally? Most importantly, how is subjective evaluation and attentional sampling weighted in the formation of the overall judgement? Through the application of eye-tracking and computational modelling across two experiments, we systematically investigated how humans valuate and attentionally sample multiple pieces of emotional evidence and how this impacts the weighing of evidence when integrating it to form a decision.

Emotional evidence is evaluated subjectively based on context.

Our results first show that subjective valuation of individual piece of evidence varies across individuals and depends on the context within which the evidence appears. Hence, in our example above, when evaluating the facial expressions of the audience, a speaker may view subtle differences in expressions as more salient in a relatively reserved audience compared to an audience who displays a wider range of emotional expressions. Such fine-tuned calibration may help individuals arrive at more adaptive decisions in their environment. While it is proposed that encoding of external evidence into an internal representation is the first step in the decision-making process (Gold & Shadlen, 2007), explicit measurement of this subjective representation, has been largely neglected in the literature. For instance, the subjective values are often either not measured (de Gardelle & Summerfield, 2011; Haberman & Whitney, 2009, 2010; Li et al., 2016) or measured outside of the context in which they are presented (Elias et al., 2017; Goldenberg et al., 2020; Sun & Chong, 2020) in multi-evidence decision-making studies. By explicitly measuring the SEV of each individual piece of evidence within the trial-context, we were able to examine how it differs from OEV and varies as a function of the overall variance across all available evidence. Relative to OEV, SEV was augmented in the low variance condition and reduced in the high variance condition. In other words, the variance of SEV was relatively stable when the variance of OEV changed, indicating that participants adopted an internal scale adjusted to the context when evaluating the intensity of external evidence.

The augmentation and reduction effects are in line with the normalization model of decision-making (Bonin et al., 2005; Carandini & Heeger, 2012; Glimcher, 2013; Heeger, 1992; Landy et al., 1995; Ohshiro et al., 2011). According to this model, the subjective value of each stimulus is normalized by the values of other stimuli available, thereby reducing redundancy and

enhancing the differences between available stimuli (Schwartz & Simoncelli, 2001). The uncovering of this pattern for emotional stimuli in the present study provides novel insight into emotion-related decision-making findings so far. For example, a recent study found that human subjects rated the average emotion of the array of faces more extreme than the objectively presented average (Goldenberg et al., 2021). The authors attributed this "emotion amplification" effect to more visual attention allocated to extreme faces which then drove the final valuation. Based on our finding, it is also possible that individual faces were already valuated as more extreme than their objective values and the participant's decision was a reflection of this subjective valuation as opposed to their attention as discussed in detail below.

Extreme evidence attracts more attention, especially when congruent with endogenous attention.

Next, we examined how evidence was sampled as a function of SEV. We found that the perceived extremity of evidence affects how it is sampled, such that extreme emotional evidence attracts more visual attention, more so when they are congruent with top-down endogenous attention. As mentioned earlier, findings from rapid decision-making tasks have implied that extreme evidence is downweighed or neglected. Due to the time constraint, it is often covert and distributed attention that is relevant in those tasks (Baek & Chong, 2020; Leib et al., 2014; Wolfe et al., 2015). Given ample time, we found that extreme evidence was not neglected. Instead, they attracted more visual attention, showing a U-shape curve when plotted against the rank order of the emotional values. Additionally, in low and high variance contexts, the faces mildest in emotion were sampled equally as the extreme faces, constituting a W-shape in Experiment 1. This pattern indicates that subjects make additional effort in examining the most extreme and most ambiguous evidence when all the pieces of evidence are either homogenous or highly different.

While Experiment 1 showed exogenously driven shifts in attention towards emotional evidence, Experiment 2 further demonstrated the effect of emotion-oriented endogenous attention towards this evidence. Our results showed that extreme emotional evidence attracted even more attention when it was the target of endogenous attention compared to when it wasn't. The finding is consistent with the cognitive conceptualization of feature-based attention, according to which endogenous attention dictates what is relevant (Summerfield & Egner, 2009), and this relevancy rendered additional salience to extreme evidence, resulting in a bottom-up and top-down interaction.

The final decision is based on the subjective valuated evidence.

Since extreme emotional stimuli receive greater exogenous and endogenous attention and because attended stimuli are weighed higher in our decisions (Krajbich et al., 2010; Orquin et al., 2021; Thomas et al., 2019), we had hypothesized that the final decision in our task would reflect greater SEV and attentional sampling. Consistent with our hypothesis, the success of the SEV model and the poor performance of the OEV model solidifies that the decisions made were very heavily based on this context-dependent subjective valuation. In contrast to our hypothesis, while participants paid more attention to evidence with higher SEV, it was not weighed higher in their overall decision. Comparing the simple averaging SEV model versus the attention-weighted SEV models, we found that more attention did not translate into higher weight in affecting the final decision, an effect that was found in both experiments. This means that even when the subjects paid much more attention to the extreme evidence congruent with their endogenous attention, they did not weigh such pieces of evidence more in their decision. Instead, all pieces of emotional evidence were weighed equally, which is the optimal decision-making strategy, in the model that best predicted their decisions. In the present task, no feedback nor reinforcer was associated with

either decision. The optimal decision-making strategy under this context was to simply average across all pieces of evidence (Van Den Berg & Ma, 2012). Hence, sufficient sampling of the evidence space coupled with an equal weight across all pieces of evidence will yield the most accurate decision about the environment. Such a decision-making strategy will facilitate individuals to interact with the environment in an adaptive manner by forming accurate judgments. Importantly, while they weigh different pieces of evidence equally, the subjective value of each piece of evidence was determined based on the presence of other pieces of evidence. Hence, the context-based calibration was intrinsic in their overall decision.

It is to be noted that while the simple averaging SEV model was the best fitting model across the entire sample, there was also existing heterogeneity. For a subsample of participants, more visual attention, reflected by the eye movement measures, further affected the weight of the corresponding evidence (i.e., greater visual attention accords that piece of evidence a larger role in the decision). For these subjects, more extreme emotional evidence attracted more visual attention and such attention was translated into higher weights of the corresponding evidence in dictating the decision.

Adaptability of the decision-making strategy.

While attentional biases may be adaptive by allowing us to sample threats and rewards in our environment, our overall decision making is not likely to be adaptive if it is heavily weighted in favor of extreme evidence that captures our attention. We show that human are generally able to avoid this and make their decision by weighing all pieces of evidence equally. However, it remains to be examined if too much attention and weight allocated to extreme emotional evidence translates into biased overall judgments in clinical conditions like anxiety or depression. For instance, anxiety is characterized by hypervigilance (Grupe & Nitschke, 2013; Power & Dalgleish,

2015), which may manifest in an increase in threat-oriented endogenous attention biasing the perceptual evidence sampling and an overweighing of threatening information. Examining whether individuals with high anxiety may allocate more attention and even more weight to negative evidence can facilitate our understanding of the related psychopathology.

Potential neural mechanisms of multi-evidence emotion-related decision making.

The dynamics between several brain regions involved in attention, value-based decision-making, and saliency processing, may play critical roles in the current task. First, the encoding of SEV is likely to involve the amygdala, the ventromedial prefrontal cortex (vmPFC) and ventral striatum. Amygdala is critical for processing salient and subjective valuated emotional information depending on the context (Anderson & Phelps, 2001; Jin et al., 2015; LeDoux, 1998). In addition, ventral striatum and vmPFC also exhibit activity correlated with the encoding of subjective values (Bartra et al., 2013; Blair, 2008; Levy & Glimcher, 2012), especially for remembering and knowing emotional valence (Mickley & Kensinger, 2008; Padoa-Schioppa & Assad, 2006; Zhang et al., 2017). Then the active sampling of emotional evidence based on SEV through eye movements may involve the lateral intraparietal area, which is known to be a hub for directing eye movements and attention and integrating visual input driven by values in decision making (Louie et al., 2011; Mohanty et al., 2009; Reynolds & Desimone, 2003). When the task-evoked endogenous attention is being manipulated, dorsolateral prefrontal cortex (dlPFC) may also be involved in initiating the top-down effect (Gray et al., 2002). Lastly, the cingulate cortex is proposed to play a key role in integrating emotion and cognition, thus it may be associated with guiding attention during evidence sampling of emotional faces based on subjective values (Mohanty et al., 2007; Shackman et al., 2011). Finally, the output may be passed down to prefrontal cortical (PFC) areas for generating binary decisions (Gold & Shadlen, 2007; Hanks &

Summerfield, 2017). Taken together, the attention network, regions encoding saliency as well as valuation involving amygdala, vmPFC, lateral intraparietal areas, amygdala, vmPFC, cingulate cortex and ventral striatum may play roles in sampling and integrating emotion evidence based on subjective valuation.

Limitations and future directions.

While this study aimed to cater toward the naturalistic setting of emotion-related perceptual decision making, the multiple stimuli used within each trial was of the same face actor, which is unrealistic. This design was chosen to ensure that the experiment was highly controlled and to address the dynamic property of facial expressions. Our findings provide foundations for further exploration such as deliberate decision making regarding a crowd of faces from multiple identities (Elias et al., 2017). Also, while the current study focused on "positive" valence as "happy", and "negative" valence as "fearful", future research is needed to test the generalizability of the findings to other categories of emotions. Lastly, in the current study, the computational modelling weights are operationalized as overt visual attention and are directly measured using eye movement data. In contrast, the weights in studies using rapid decision are mostly statistically inferred parameter estimates. The two types of operationalization of weights may not correspond to each other and therefore lead to different conclusions. While we think it is unlikely that people would weigh a piece of evidence highly when not paying much attention to it, future research may help unpack what psychological processes (if not attention) do the parameter estimates best account for.

Concluding remarks.

Decision making regarding emotional information has been a long-standing interest in affective and cognitive sciences. Findings from the present study help elucidate how humans valuate, sample, and weigh multiple pieces of emotional evidence to form an overall judgment.

The psychological and computational mechanisms uncovered in the present study also shed light on how emotion-related decision making can be biased in people with emotion-related psychopathology.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122-131. https://doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features outside the Focus of Attention. *Psychological science*, *19*(4), 392-398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature (London)*, *411*(6835), 305-309. https://doi.org/10.1038/35077083

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological science*, *12*(2), 157-162. https://doi.org/10.1111/1467-9280.00327

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision making*, *3*(5), 396-403.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & psychophysics*, *55*(5), 485-496.

Baek, J., & Chong, S. C. (2020). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic bulletin & review*, *27*(4), 602-606. https://doi.org/10.3758/s13423-020-01718-7

Barrett, L. F. (2006). Solving the emotion paradox: categorization and the experience of emotion. *Personality and social psychology review*, *10*(1), 20-46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature reviews neuroscience*, *16*(7), 419-429.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage (Orlando, Fla.)*, *76*, 412-427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Bhatia, S. (2014). Confirmatory search and asymmetric dominance. *Journal of Behavioral Decision Making*, *27*(5), 468-476.

Blair, R. J. R. (2008). The amygdala and ventromedial prefrontal cortex: functional contributions and dysfunction in psychopathy. *Philosophical transactions. Biological sciences*, *363*(1503), 2557-2565. https://doi.org/10.1098/rstb.2008.0027

Bonin, V., Mante, V., & Carandini, M. (2005). The Suppressive Field of Neurons in Lateral Geniculate Nucleus. *The Journal of neuroscience*, *25*(47), 10844-10856. https://doi.org/10.1523/JNEUROSCI.3562-05.2005

Buck, R. (1980). Nonverbal behavior and the theory of emotion: The facial feedback hypothesis. *Journal of personality and social psychology*, *38*(5), 811-824. https://doi.org/10.1037/0022-3514.38.5.811

Buck, R. (1988). *Human motivation and emotion*. John Wiley & Sons.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and emotion*, *30*(6), 1081-1106. https://doi.org/10.1080/02699931.2015.1049124

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51-62.

Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature neuroscience*, *4*(7), 673-674. https://doi.org/10.1038/89436

Chan, C. Y. H., Chan, A. B., Lee, T. M. C., & Hsiao, J. H. (2018). Eye-movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic bulletin & review*, *25*(6), 2200-2207. https://doi.org/10.3758/s13423-017-1419-0

Cho, V. Y., Hsiao, J. H., Chan, A. B., Ngo, H., King, N. M., & Anthonappa, R. (2022a). Understanding children's attention to dental caries through eye-tracking. *Caries Research*.

Cho, V. Y., Hsiao, J. H., Chan, A. B., Ngo, H. C., King, N. M., & Anthonappa, R. P. (2022b). Understanding children's attention to traumatic dental injuries using eye-tracking. *Dental traumatology*.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393-404.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *Journal of vision (Charlottesville, Va.)*, *14*(11), 8-8. https://doi.org/10.1167/14.11.8

Chuk, T., Chan, A. B., & Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling. *Vision research (Oxford)*, *141*, 204-216. https://doi.org/10.1016/j.visres.2017.03.010

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2019). Eye movement analysis with switching hidden Markov models. *Behavior research methods*, *52*(3), 1026-1043. https://doi.org/10.3758/s13428-019-01298-y

Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, *169*, 102-117. https://doi.org/10.1016/j.cognition.2017.08.003

Clark, C. C., & Lawless, H. T. (1994). Limiting response alternatives in time-intensity scaling: an examination of the halo-dumping effect. *Chemical senses*, *19*(6), 583-594. https://doi.org/10.1093/chemse/19.6.583

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision research (Oxford)*, *37*(22), 3181-3192. https://doi.org/10.1016/S0042-6989(97)00133-8

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences - PNAS*, *108*(32), 13341-13346. https://doi.org/10.1073/pnas.1104517108

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble Perception of Dynamic Emotional Groups. *Psychological science*, *28*(2), 193-203. https://doi.org/10.1177/0956797616678188

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The Outlier Paradox: The Role of Iterative Ensemble Coding in Discounting Outliers. *Journal of experimental psychology. Human perception and performance*, *46*(11), 1267-1279. https://doi.org/10.1037/xhp0000857

Feldman Barrett, L. (2011). Constructing emotion. *Psihologijske teme*, *20*(3), 359-380.

Feldman, L. A. (1995). Variations in the Circumplex Structure of Mood. *Personality & social psychology bulletin*, *21*(8), 806-817. https://doi.org/10.1177/0146167295218003

Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, *6*(1).

Glasgow, S., Imbriano, G., Jin, J., Zhang, X., & Mohanty, A. (2022). Threat and uncertainty in the face of perceptual decision-making in anxiety. *Journal of Psychopathology and Clinical Science*, *131*(3), 265.

Glimcher, P. W. (2013). *Neuroeconomics: Decision Making and the Brain* (E. Fehr, Ed.). Academic Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*(1), 535-574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Goldenberg, A., Sweeny, T. D., Shpigel, E., & Gross, J. J. (2020). Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *Journal of Experimental Psychology: General*, *149*(3), 445.

Goldenberg, A., Weisz, E., Sweeny, T. D., Cikara, M., & Gross, J. J. (2021). The crowd-emotion-amplification effect. *Psychological science*, *32*(3), 437-450.

Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences*, *99*(6), 4115-4120. https://doi.org/10.1073/pnas.062381899

Grebitus, C., Roosen, J., & Seitz, C. C. (2015). Visual Attention and Choice: A Behavioral Economics Perspective on Food Decisions. *Journal of Agricultural & Food Industrial Organization*, *13*(1), 73-81. https://doi.org/doi:10.1515/jafio-2015-0017

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature reviews. Neuroscience*, *14*(7), 488-501. https://doi.org/10.1038/nrn3524

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of vision (Charlottesville, Va.)*, *9*(11), 1.1-1. https://doi.org/10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of vision (Charlottesville, Va.)*, *15*(4), 16-16. https://doi.org/10.1167/15.4.16

Haberman, J., & Whitney, D. (2009). Seeing the Mean: Ensemble Coding for Sets of Faces. *Journal of experimental psychology. Human perception and performance*, *35*(3), 718-734. https://doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, perception & psychophysics*, *72*(7), 1825-1838. https://doi.org/10.3758/APP.72.7.1825

Hahn, S., & Gronlund, S. D. (2007). Top-down guidance in visual search for facial expressions. *Psychonomic Bulletin & Review*, *14*(1), 159-165.

Hanks, T. D., & Summerfield, C. (2017). Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron (Cambridge, Mass.)*, *93*(1), 15-31. https://doi.org/10.1016/j.neuron.2016.12.003

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual neuroscience*, *9*(2), 181-197. https://doi.org/10.1017/S0952523800009640

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in cognitive sciences*, *13*(12), 517-523.

Huber, P. J. (2004). *Robust statistics* (Vol. 523). John Wiley & Sons.

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of vision (Charlottesville, Va.)*, *15*(4), 5-5. https://doi.org/10.1167/15.4.5

Imbriano, G., Sussman, T. J., Jin, J., & Mohanty, A. (2020). The role of imagery in threat-related perceptual decision making. *Emotion*, *20*(8), 1495.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489-1506.

Ji, L., Chen, W., & Fu, X. (2014). Different Roles of Foveal and Extrafoveal Vision in Ensemble Representation for Facial Expressions.*Lecture Notes in Computer Science* Cham.

Jin, J., Zelano, C., Gottfried, J. A., & Mohanty, A. (2015). Human Amygdala Represents the Complete Spectrum of Subjective Valence. *The Journal of neuroscience*, *35*(45), 15145-15156. https://doi.org/10.1523/JNEUROSCI.2450-15.2015

Kaanders, P., Sepulveda, P., Folke, T., Ortoleva, P., & De Martino, B. (2022). Humans actively sample evidence to support prior beliefs. *Elife*, *11*, e71768.

Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt Visual Attention as a Causal Factor of Perceptual Awareness. *PLOS ONE*, *6*(7), e22614. https://doi.org/10.1371/journal.pone.0022614

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292-1298. https://doi.org/10.1038/nn.2635

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences - PNAS*, *108*(33), 13852-13857. https://doi.org/10.1073/pnas.1101328108

Kulke, L. (2019). Neural Mechanisms of Overt Attention Shifts to Emotional Faces. *Neuroscience*, *418*, 59-68. https://doi.org/10.1016/j.neuroscience.2019.08.023

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision research (Oxford)*, *35*(3), 389-412. https://doi.org/10.1016/0042-6989(94)00176-M

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated Attention: Affect, Activation, and Action. In (pp. 125-164). Psychology Press. https://doi.org/10.4324/9780203726457-14

Lawless, H. T., & Clark, C. C. (1992). Psychological biases in time-intensity scaling. *Food technology (USA)*.

LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster.

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: a viewpoint-invariant mechanism to represent average crowd identity. *Journal of vision (Charlottesville, Va.)*, *14*(8), 26-26. https://doi.org/10.1167/14.8.26

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, *22*(6), 1027-1038. https://doi.org/10.1016/j.conb.2012.06.001

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in psychology*, *7*, 1332-1332. https://doi.org/10.3389/fpsyg.2016.01332

Li, V., Herce Castanon, S., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLoS computational biology*, *13*(8), e1005723-e1005723. https://doi.org/10.1371/journal.pcbi.1005723

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *The Journal of neuroscience*, *31*(29), 10627-10639. https://doi.org/10.1523/JNEUROSCI.1237-11.2011

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Veerdonk, E. V. d. (2008). Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. *Journal of personality and social psychology*, *94*(3), 365-381. https://doi.org/10.1037/0022-3514.94.3.365

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences - PNAS*, *111*(21), 7873-7878. https://doi.org/10.1073/pnas.1308674111

Mickley, K. R., & Kensinger, E. A. (2008). Emotional valence influences the neural correlates associated with remembering and knowing. *Cognitive, affective, & behavioral neuroscience*, *8*(2), 143-152. https://doi.org/10.3758/CABN.8.2.143

Mogg, K., & Bradley, B. P. (1999). Orienting of Attention to Threatening Facial Expressions Presented under Conditions of Restricted Awareness. *Cognition and emotion*, *13*(6), 713-740. https://doi.org/10.1080/026999399379050

Mohanty, A., Egner, T., Monti, J. M., & Mesulam, M.-M. (2009). Search for a threatening target triggers limbic guidance of spatial attention. *Journal of Neuroscience*, *29*(34), 10563-10572.

Mohanty, A., Engels, A. S., Herrington, J. D., Heller, W., Ringo Ho, M. H., Banich, M. T., Webb, A. G., Warren, S. L., & Miller, G. A. (2007). Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. *Psychophysiology*, *44*(3), 343-351.

Mohanty, A., Jin, F., & Sussman, T. (2023). What Do We Know About Threat-Related Perceptual Decision Making? *Current Directions in Psychological Science*, *0*(0), 09637214221129795. https://doi.org/10.1177/09637214221129795

Mohanty, A., & Sussman, T. J. (2013). Top-down modulation of attention by emotion. In (Vol. 7, pp. 102): Frontiers Media SA.

Nummenmaa, L., Hari, R., Hietanen, J. K., & Glerean, E. (2018). Maps of subjective feelings. *Proceedings of the National Academy of Sciences*, *115*(37), 9198-9203.

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye Movement Assessment of Selective Attentional Capture by Emotional Pictures. *Emotion (Washington, D.C.)*, *6*(2), 257-268. https://doi.org/10.1037/1528-3542.6.2.257

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature neuroscience*, *14*(6), 775-782. https://doi.org/10.1038/nn.2815

Orquin, J. L., Lahm, E. S., & Stojić, H. (2021). The visual environment and attention in decision making. *Psychological Bulletin*, *147*(6), 597.

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta psychologica*, *144*(1), 190-206.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223-226. https://doi.org/10.1038/nature04676

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, *4*(7), 739-744. https://doi.org/10.1038/89532

Pärnamets, P., Johansson, R., Gidlöf, K., & Wallin, A. (2016). How information availability interacts with visual attention during judgment and decision tasks. *Journal of Behavioral Decision Making*, *29*(2-3), 218-231.

Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *NeuroImage (Orlando, Fla.)*, *47*(4), 1873-1883. https://doi.org/10.1016/j.neuroimage.2009.03.084

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, *17*(3), 715-734. https://doi.org/10.1017/S0954579405050340

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological review*, *78*(5), 391.

Power, M., & Dalgleish, T. (2015). *Cognition and Emotion: From order to disorder*. Taylor and Francis. https://doi.org/10.4324/9781315708744

Rahn, J., Jaudas, A., & Achtziger, A. (2016). To plan or not to plan—mindset effects on visual attention in decision making. *Journal of Neuroscience, Psychology, and Economics*, *9*(2), 109-120. https://doi.org/https://doi.org/10.1037/npe0000056

Reynolds, J. H., & Desimone, R. (2003). Interacting Roles of Attention and Visual Salience in V4. *Neuron (Cambridge, Mass.)*, *37*(5), 853-863. https://doi.org/10.1016/S0896-6273(03)00097-7

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature neuroscience*, *4*(8), 819-825. https://doi.org/10.1038/90526

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *eLife*, *9*. https://doi.org/10.7554/eLife.60705

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154-167.

Smith, P. L., & Ratcliff, R. (2009). An Integrated Theory of Attention and Decision Making in Visual Signal Detection. *Psychological review*, *116*(2), 283-317. https://doi.org/10.1037/a0015156

Spektor, M. S., Bhatia, S., & Gluth, S. (2021). The elusiveness of context effects in decision making. *Trends in Cognitive Sciences*, *25*(10), 843-854.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*(9), 403-409.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311-1314.

Sun, J., & Chong, S. C. (2020). Power of Averaging: Noise Reduction by Ensemble Coding of Multiple Faces. *Journal of experimental psychology. General*, *149*(3), 550-563. https://doi.org/10.1037/xge0000667

Sussman, T. J., Jin, J., & Mohanty, A. (2016). Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biological psychology*, *121*, 160-172.

Sussman, T. J., Szekely, A., Hajcak, G., & Mohanty, A. (2016). It's all in the anticipation: How perception of threat is enhanced in anxiety. *Emotion*, *16*(3), 320.

Sussman, T. J., Weinberg, A., Szekely, A., Hajcak, G., & Mohanty, A. (2017). Here comes trouble: Prestimulus brain activity predicts enhanced perception of threat. *Cerebral Cortex*, *27*(4), 2695-2707.

Taylor, J. G., & Fragopanagos, N. F. (2005). The interaction of attention and emotion. *Neural networks*, *18*(4), 353-369.

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625-635.

Van Den Berg, R., & Ma, W. J. (2012). Robust averaging during perceptual judgment is not optimal. *Proceedings of the National Academy of Sciences*, *109*(13), E736-E736.

Vandormael, H., Herce Castañón, S., Balaguer, J., Li, V., & Summerfield, C. (2017). Robust sampling of decision information during perceptual choice. *Proceedings of the National Academy of Sciences - PNAS*, *114*(10), 2771-2776. https://doi.org/10.1073/pnas.1613950114

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision research (Oxford)*, *32*(5), 931-941. https://doi.org/10.1016/0042-6989(92)90036-I

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision research (Oxford)*, *29*(1), 47-59. https://doi.org/10.1016/0042-6989(89)90173-9

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence. *Journal of personality and social psychology*, *76*(5), 820-838. https://doi.org/10.1037/0022-3514.76.5.820

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual review of psychology*, *69*(1), 105-129. https://doi.org/10.1146/annurev-psych-010416-044232

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of vision (Charlottesville, Va.)*, *15*(4), 11-11. https://doi.org/10.1167/15.4.11

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *Journal of Neuroscience*, *28*(10), 2667-2679.

Zhang, J., Chan, A. B., Lau, E. Y. Y., & Hsiao, J. H. (2019). Individuals with insomnia misrecognize angry faces as fearful faces while missing the eyes: An eye-tracking study. *Sleep*, *42*(2), zsy220.

Zhang, Z., Fanning, J., Ehrlich, D. B., Chen, W., Lee, D., & Levy, I. (2017). Distributed neural representation of saliency controlled value and category during anticipation of rewards and punishments. *Nature communications*, *8*(1), 1907-1914. https://doi.org/10.1038/s41467-017-02080-4

Figures

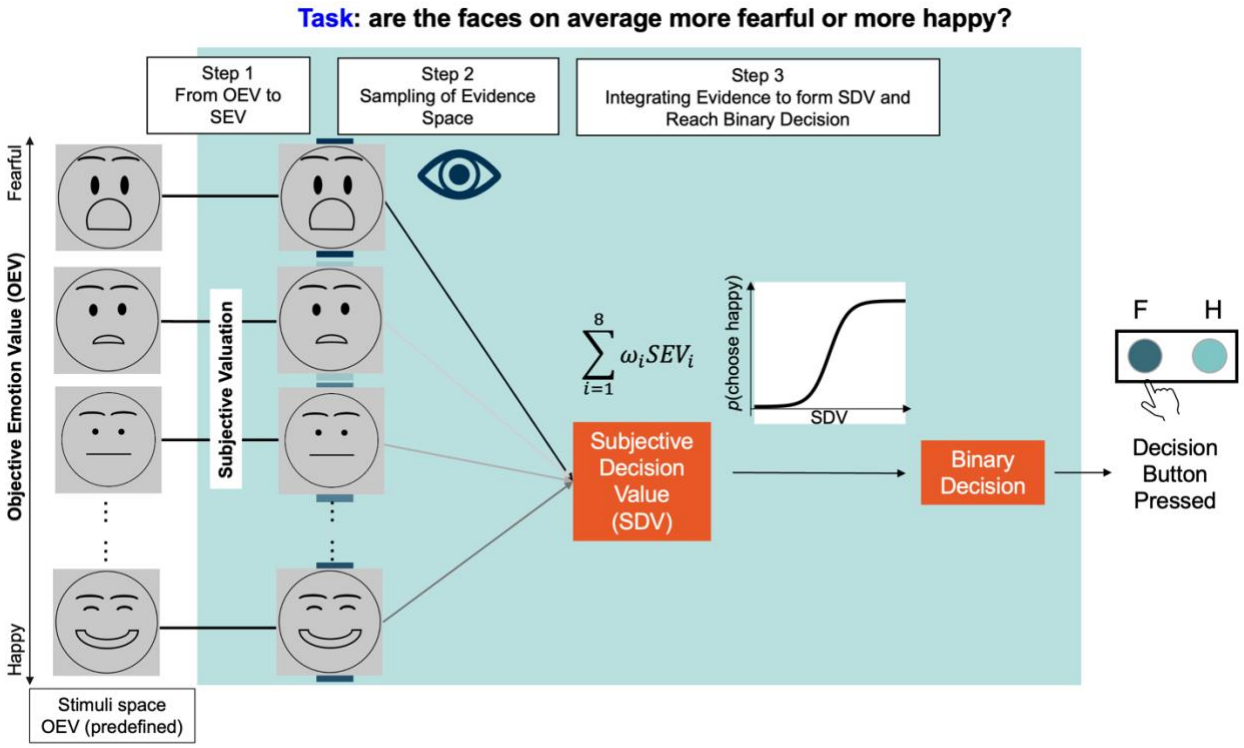

Figure 1. **The multi-evidence emotion-related perceptual decision-making process.** The process is dissected into the following three components: (1) mapping from the external evidence space (each evidence has an objective emotion value, OEV) to the internal space (each piece of evidence has a subjective emotion value, SEV); (2) sampling the evidence space, which can be captured by visual attention; then (3) integration of these pieces of evidence (e.g., weighted average) to form a Subjective Decision Value (SDV), which is then passed through a softmax function to form a binary decision (with certain level of stochasticity). The subject will announce his or her decision by pressing a button).

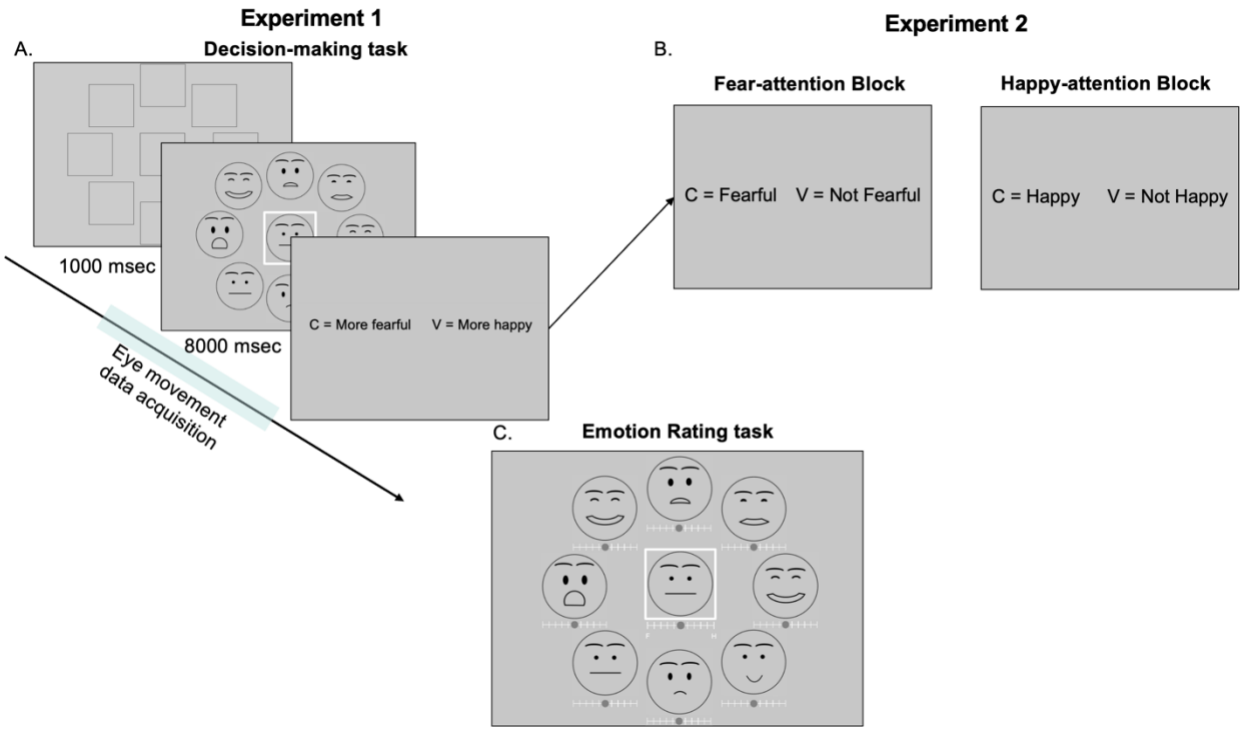

Figure 2. **Experiment 1 & 2 tasks. (A)** Trial timeline of the decision-making task. After empty placeholders, 8 faces were shown for 8000 msec, during which eye movements were recorded. Subjects made a binary decision after the faces were gone. In Experiment 1, subjects decided whether the faces were more fearful or more happy. **(B)** In Experiment 2, subjects were asked to make a "Fearful or Not Fearful" decision in the Fear-attention block, and a "Happy or Not Happy" decision in the Happy-attention block. **(C)** Sample trial for the subjective emotion rating task. Instead of making a binary decision, subjects rated each individual face on a scale of Fearful to Happy without time constraints. In both **A** and **C**, the center face was that trial's particular face model's face with the OEV of 50. Subjects were instructed to reference that face when making a decision and rating. The subjective emotion rating task was administered in both Experiment 1 and 2.

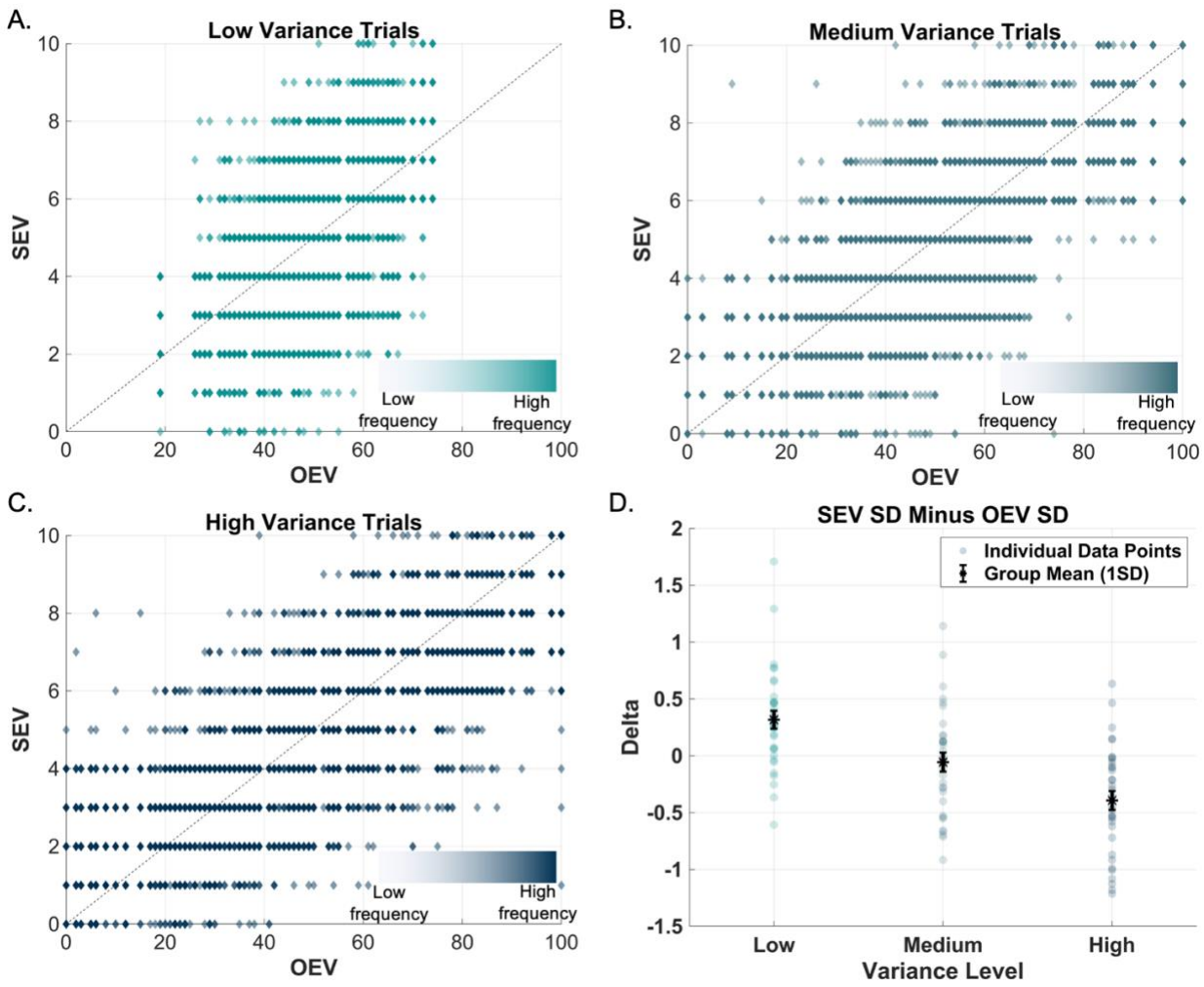

Figure 3. **Experiment 1 Behavioral Rating and Decision-making Results.** Associations between subject-rated SEV and experimenter-defined OEV in **(A)** Low (**B)** Medium and (**C)** High variance trials. **(D)** SD of SEV subtracting SD of OEV for low, medium and high variance trials where circles indicate individual subject's data and asterisk indicates the sample mean. Error bars represent ±1 SD. For **A** to **D**, the transparency of the dots represents data frequency, with low transparency being higher frequency.

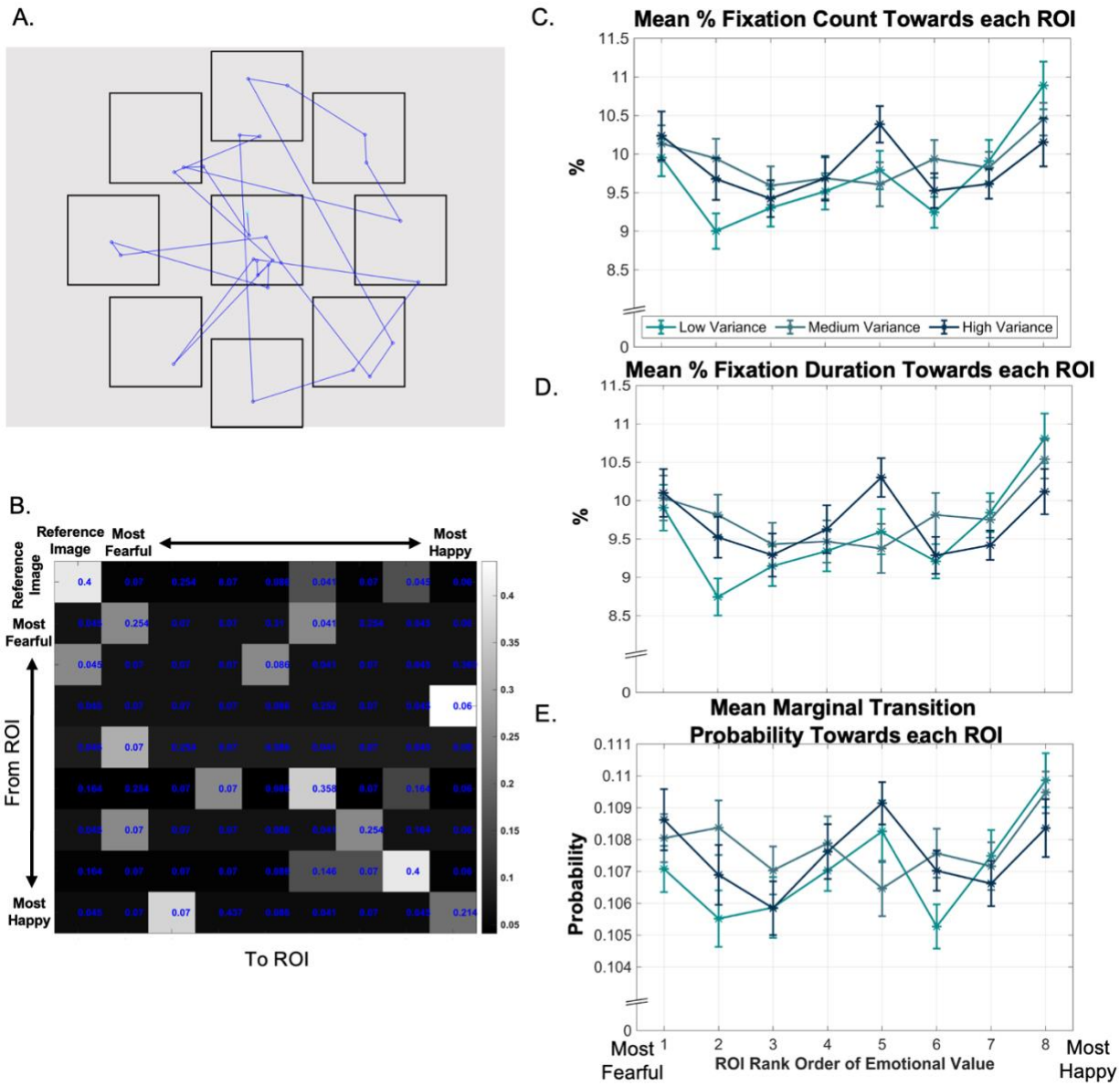

***Figure 4*. Experiment 1 Eye Movement Results. (A)** Sample subject's trajectory of eye movement during the 8000 msec in one trial, where the cyan dot indicates the first fixation of that trial. **(B)** Sample EMHMM output of transition matrix, where the transition probabilities have been reordered according to SEV and OEV, so that the top and left is the ROI pertaining to the central reference image, next to which is the most fearful ROI of that trial while the bottom and right is the happiest ROI of that trial. Each cell indicates the probability of transitioning from a given ROI to another ROI. **(C)** The mean fixation count (%) and **(D)** fixation duration (%) and **(E)** marginal transition probability towards the eight evidence-bearing images in all subjects across the three variance levels; the images have been reordered so 1 indicates the most fearful ROI of all trials and 8 is the happiest ROI of all trials. Error bars represent ±1 SD.

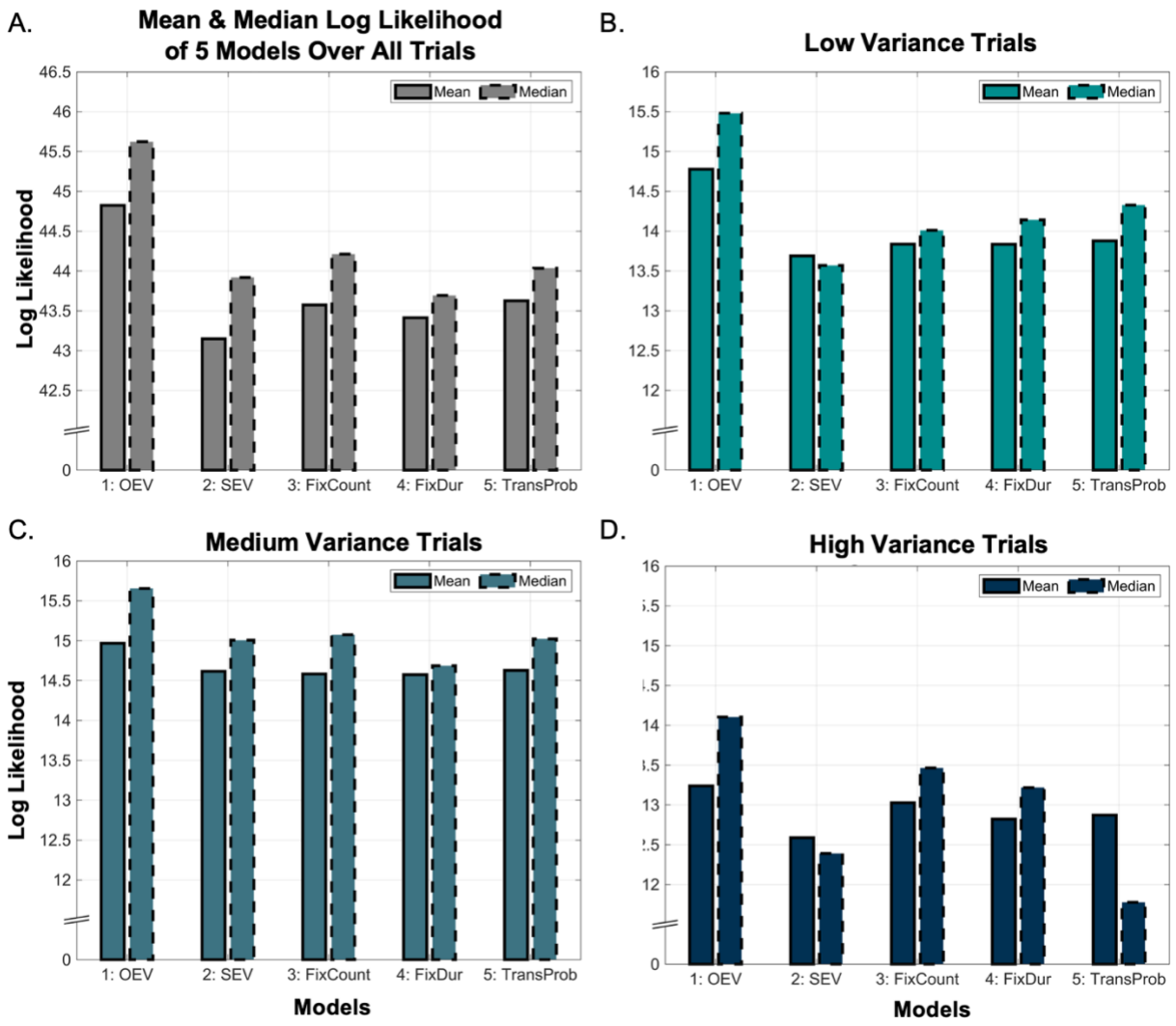

Figure 5. **Experiment 1 Model comparison results. (A)** Mean and median of each model's minimized log likelihood over all trials. The results show that Model 2, where the SDV was calculated as a simple average of all SEV had the lowest mean log likelihood. Model 4, where the SDV was calculated as a weighted average of SEV with its corresponding fixation duration proportion had the lowest median log likelihood. Mean and median of each model's minimized log likelihood for trials of **(B)** low variance, **(C)** medium variance, and **(D)** high variance.

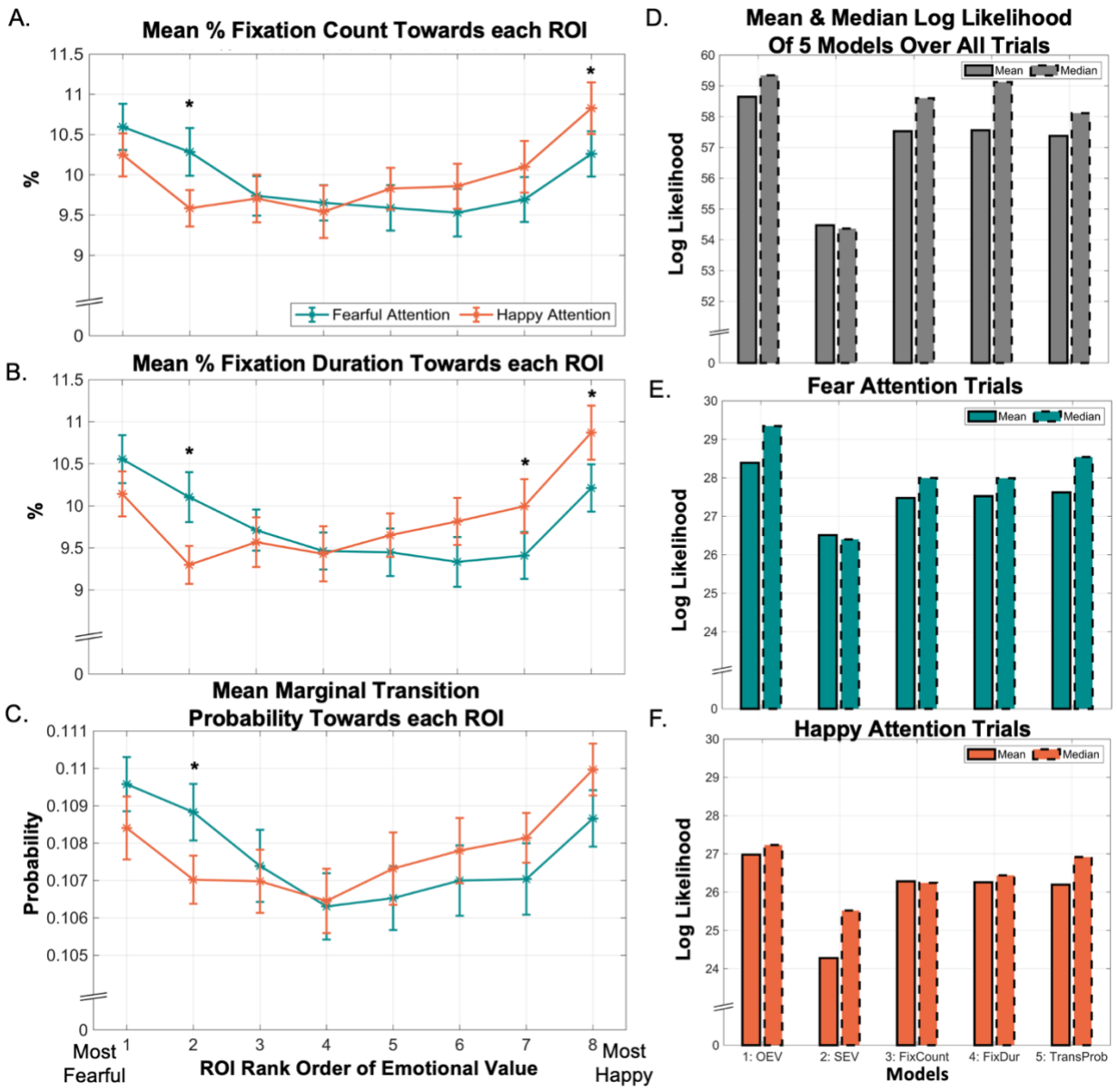

Figure 6. **Experiment 2 Eye-movement & Computational Modelling Results.** The mean **(A)** fixation count (%), **(B)** fixation duration (%), and **(C)** Marginal transition probability towards the eight evidence-bearing images in all subjects across the two endogenous attention manipulation conditions. Green refers to (F)earful-attention and orange refers to (H)appy-attention (* marks $p < .05$). Mean and median of each model's minimized log likelihood across **(D)** all trials **(E)** (F)earful-attention trials and **(F)** (H)appy-attention trials.