Normal observers show no evidence for blindsight in facial emotion perception

Sivananda Rajananda[1], Jeanette Zhu ⓘ [2,†] and Megan A. K. Peters ⓘ [1,3,*,‡]

[1]Department of Bioengineering, University of California Riverside, Riverside, CA 92521, USA; [2]Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA; [3]Department of Cognitive Science, University of California Irvine, Irvine, CA 92697, USA

*Correspondence address. Department of Cognitive Sciences, University of California, Irvine, 2201 Social & Behavioral Sciences Gateway Building, Irvine, CA 92697, USA. E-mail: megan.peters@uci.edu

[†]Jeanette Zhu, https://orcid.org/0000-0002-4087-1279

[‡]Megan A. K. Peters, https://orcid.org/0000-0002-0248-0816

Abstract

Some researchers have argued that normal human observers can exhibit "blindsight-like" behavior: the ability to discriminate or identify a stimulus without being aware of it. However, we recently used a bias-free task to show that what looks like blindsight may in fact be an artifact of typical experimental paradigms' susceptibility to response bias. While those findings challenge previous reports of blindsight in normal observers, they do not rule out the possibility that different stimuli or techniques could still reveal perception without awareness. One intriguing candidate is emotion processing, since processing of emotional stimuli (e.g. fearful/happy faces) has been reported to potentially bypass conscious visual circuits. Here we used the bias-free blindsight paradigm to investigate whether emotion processing might reveal "featural blindsight," i.e. ability to identify a face's emotion without introspective access to the task-relevant features that led to the discrimination decision. However, we saw no evidence for emotion processing "featural blindsight": as before, whenever participants could identify a face's emotion they displayed introspective access to the task-relevant features, matching predictions of a Bayesian ideal observer. These results add to the growing body of evidence that perceptual discrimination ability without introspective access may not be possible for neurologically intact observers.

Keywords: blindsight; featural blindsight; emotion perception; fear detection; Bayesian ideal observer; metacognition; consciousness

Introduction

The neurological condition of blindsight has fascinated consciousness researchers since its discovery (Weiskrantz, 1986, 1996; Azzopardi and Cowey 1997, 1998); in this rare condition, patients with damage to primary visual cortex can directly discriminate some aspects of visual stimuli but report no conscious visual awareness of them. Thus, to experimentally use this dissociation to isolate conscious awareness of a stimulus from the potential confound of signal processing capacity (Lau, 2008; Morales et al. 2015;

Giles et al. 2016; Peters et al. 2017b), we must find a way to induce blindsight-like behavior in neurologically intact observers. Note that, we aim to induce absolute blindsight, where consciousness awareness is abolished completely, as opposed to relative blindsight, where performance is matched while confidence differs (Lau and Passingham 2006; Balsdon and Azzopardi 2015; Peters and Lau 2015; Peters et al. 2017a; Knotts et al., 2018). Successful induction of such "performance without awareness" would pave the way for computational and neuroimaging studies seeking to

identify the neural correlates of consciousness (Rees *et al.* 2002; Block, 2005; Tononi and Koch 2008; Aru *et al.* 2012).

To measure unconscious perception, researchers have generally been divided on two approaches: those relying on the *objective threshold*, and those relying on the *subjective threshold* as the relevant criterion for awareness of a stimulus (c.f. Snodgrass and Shevrin 2006). Objective threshold advocates suggest that performance at chance ($d' = 0$) is the threshold at which the subject becomes conscious of the stimulus (i.e. the subject is conscious of the stimulus if $d' > 0$). Subjective threshold advocates suggest that the threshold at which a stimulus becomes conscious is the same threshold at which they are able to report it (i.e. only if the subject reports that they saw the stimulus are they conscious of it, meaning $d' > 0$ does not necessarily indicate consciousness of the stimulus).

Researchers relying on the subjective threshold definition for stimulus awareness have often attempted to induce blindsight-like behavior in normal observers, typically using visual masking or other similar techniques. Unfortunately, most of these attempts are susceptible to response bias confounds (Eriksen, 1960; Merikle *et al.* 2001; Ramsøy and Overgaard 2004; Hannula *et al.* 2005; Charles *et al.* 2013, 2014; Lloyd *et al.* 2013; Jachs *et al.* 2015; Peters *et al.* 2016, 2017a). Further, past attempts to remedy response bias concerns (Kolb and Braun 1995; Kunimoto *et al.* 2001) have met with their own conceptual or replicability challenges (Morgan *et al.* 1997; Galvin *et al.* 2003; Robichaud and Stelmach 2003; Evans and Azzopardi 2007). Recently, we used visual masking in a bias-free paradigm to demonstrate that it may not be possible to induce blindsight in normal observers when response bias confounds are controlled for (Peters and Lau 2015); follow-up studies also demonstrated that several other masking techniques commonly assumed to dissociate objective and subjective processing may similarly fail to produce conditions under which blindsight could be induced (Knotts *et al.*, 2018).

Yet these failures do not unequivocally prove that blindsight-like behavior *cannot* be induced in normal observers. Although we failed to demonstrate that normal observers can discriminate a stimulus yet be completely unaware of its *overall presence* (Peters and Lau 2015), it may be possible for an observer to be able to discriminate a stimulus above chance while being unaware of its *task-relevant properties*. This might occur especially for stimuli that could possibly bypass "conscious" visual processing areas to activate other task-relevant circuitry, such as observations of amygdala reactivity in "unconscious" face emotion processing (Pessoa and Adolphs 2010; Watanabe and Haruno 2015; Diano *et al.* 2016; Khalid and Ansorge 2017). In such emotion processing "featural blindsight," the observer would be aware of the face itself and be able to correctly identify its emotion (e.g. happy versus fearful), but would report no subjective or introspective access to the *task-relevant features*, evinced by no introspective confidence in their choices. Indeed, it has recently been reported that processing of emotional stimuli, especially fearful stimuli, may be a powerful candidate for inducing blindsight-like behavior (Vieira *et al.* 2017).

Therefore, here we used emotional face stimuli matched on all low-level properties (e.g. luminance, contrast, spatial frequency) to investigate whether normal observers can identify facial emotions without being aware of their ability to do so. Importantly, we controlled for response bias using the same bias-free confidence paradigm (Mamassian 2020) previously shown to demonstrate optimal introspective access using low-level stimuli (i.e. masked Gabor patches) (Peters and Lau 2015). If we observe emotion-processing "featural blindsight" using

this paradigm, the results would suggest that the previous failure to demonstrate blindsight in normal observers was due to the unsuitability of impoverished, low-level stimuli to dissociate objective versus subjective thresholds (Peters and Lau 2015)—and further, that emotion processing may provide an ideal candidate for experimentally isolating subjective awareness from objective task performance capacity in future studies seeking the neural or computational correlates of consciousness.

Materials and Methods
Behavioral experiment

Participants
Participants for this experiment were recruited from Amazon Mechanical Turk. The task was launched for 40 people; 33 of these participants completed the experiment. Those who completed the experiment were paid $2. Participants were also incentivized to perform well with an extra $1 which was awarded if they performed better than the previous participant (incentive structure is described under Procedure section). Of the 33 people who completed the experiment, 2 had incomplete data, and 2 had >20% of trials which met the exclusion criteria (see below); data from the remaining 29 participants was included in all further analyses. Informed consent was obtained before the start of the experiment, and all procedures were approved by the University of California, Riverside Institutional Review Board and made in accordance with the Declaration of Helsinki.

Stimuli
Stimuli consisted of four sets of male and four sets of female faces displaying happiness ("happy" faces), fear ("fearful" faces), or no emotion ("neutral" faces). The faces were selected from the NimStim set of facial expressions (Tottenham *et al.* 2009), which contains images of 43 (18 male, 25 female) people displaying various emotions and neutral expressions. In an earlier experiment (also via Amazon Mechanical Turk), 20 participants viewed 15 female faces and 13 male faces displaying happiness and fear at maximal intensity, and rated the intensity of the emotion shown by each face on a scale of 0–100, where 0 is "completely neutral" and 100 is the "most possible fear/happiness that a person could exhibit." We selected the four highest-rated faces from each gender for which the difference in the emotional intensity ratings was not significantly different for "happy" versus "fearful" images.

The selected four male and four female faces were then morphed between the "neutral" image and the "happy" and "fearful" images for each person, to create a set of eight morphs from 0% emotional intensity to 100% emotional intensity for each emotion. We then matched the low-level image properties of the morphs using the SHINE toolbox (Willenbockel *et al.* 2010): image luminance histograms of the photos were matched through SHINE toolbox's histMatch function with the structural similarity also optimized through structural similarity (SSIM) gradient ascent (20 iterations), and the amplitude spectrum were matched across images through the specMatch function.

Procedure
On each trial, subjects viewed two faces presented sequentially, one after the other. Following the 2-interval forced-choice (2IFC) confidence-betting paradigm (Mamassian 2020) used by Peters and Lau (2015), one interval showed the "neutral" face [Emotion-Absent interval (EA)] and the other showed the same

person's face but with some degree of emotion, selected from the morphed images [Emotion-Present interval (EP)]. The EA and EP intervals were counterbalanced randomly such that the subject would have a 50% of being correct if they bet on either interval. Subjects were asked to determine the emotion of each face (happy/fearful). Subjects were not explicitly informed that one of the intervals had no emotion [which allowed us to compare betting on intervals where there is evidence (EP) versus where there is not (EA)]. Thus, if we find that subjects bet on both EA and EP intervals equally, we can conclude that the metacognitive or introspective confidence—i.e., the awareness of task-relevant stimulus features—for EP intervals is equivalent to that in EA intervals, indicating that subjects had no introspective access to these task-relevant features in the EP intervals. Subjects were also instructed that the two intervals' emotions were independently chosen, such that any combination of emotion pairs was possible: happy/happy, happy/fearful, fearful/happy, and fearful/fearful. In addition to identifying the emotion, subjects were asked to bet on which emotional identification they felt more confident in. To incentivize their performance, they were told that for each emotion that they identified correctly, they would get a point. They were also told that if they bet on the easier interval, i.e., the one they thought they were more likely to get correct, they would also get a point. These points were then used to determine if they earned the bonus payment.

Each trial began with a fixation point (1000 ms) followed by the first interval's face presented centrally (size dynamically scaled based on window size to occupy ∼60% of screen height) (500 ms), a second fixation point (1000 ms), and the second interval's face presented centrally (500 ms) (Fig. 1). The order of EP and EA intervals was counterbalanced across trials, and the emotion to be presented in the EP interval was chosen pseudorandomly such that 50% of trials presented a fearful face and 50% of trials presented a happy face in the EP interval. Each of the two faces used on each trial was chosen randomly from the eight possible face sets, and both faces within a trial were from different sets. For the EP interval in each trial, the emotional intensity of the face (whether happy or fearful) was pseudorandomly set to be 5%, 15%, or 25% for 32 trials each, or 75% for 16 trials (75% is very easy).

After the second interval's face had disappeared, subjects indicated which interval they felt more confident in (Interval Question), and then the emotion decision (Emotion Questions) for the first and second face, respectively. To indicate their answers, subjects entered their "emotion" choices using the "H"

key (happy) and "F" key (fearful) on their keyboard using their right hand, and their choices about which decision to "bet" on using the "1" and "2" keys using their left hand. The next trial began as soon as subjects had pressed a key or after 6 s had elapsed, whichever happened sooner. As described above, only one of the intervals contained an emotional face, with the other containing a neutral face—despite the fact that subjects were told both intervals contained emotional faces.

Before beginning the main experiment, participants completed four easy practice trials and four hard practice trials. Practice trials were identical to trials in the main experiment, except that the faces presented in both intervals were presented for 1000 ms instead of 500 ms, and *both* faces had an emotion present to reinforce the instructions given to subjects to judge the emotion in *both* intervals. Although we could have informed them that only one face contained emotion, we wanted subjects to be equally motivated in both intervals rather than the potential for seeing a clear emotion in the first interval and then not paying attention to the second interval. The emotion intensity in the easy practice trials was 100%, while the emotion intensity in the hard practice trials was 25%. Prior to the beginning of the hard practice trials, participants were told that some emotions would be hard to distinguish, and if they were unsure, they should just make their best guess. After the 8 practice trials, the participants completed 112 trials in the main experiment.

Participants were given two breaks: one after trial 37, and another after trial 74. Each break lasted 1 min, although participants were encouraged to take more time if they needed. In total, the experiment lasted ∼30 min.

Analysis

To determine the objective (Type 1) performance capacity for each participant, we calculated the percent of EP trials in which the emotion was correctly identified separately for each emotional intensity level (Emotion Questions). This produced four "% correct emotion discrimination" values for each subject. Following Peters and Lau (2015), we also calculated the confidence/introspective ability of each observer (Type 2) as the percent of trials the subject "bet" on the EP interval at each of the four emotional intensity levels (Interval Question; "% bet on EP"). Trials were excluded from the "% bet on EP" analysis if reaction times to any question (Emotion or Interval Questions) in the trial exceeded 4000 ms, or if the subject failed to respond to all questions in the trial. Likewise, trials were excluded from the "% correct" analysis if reaction time to the Interval Question or the Emotion Question for the EP interval exceeded 4000 ms, or if
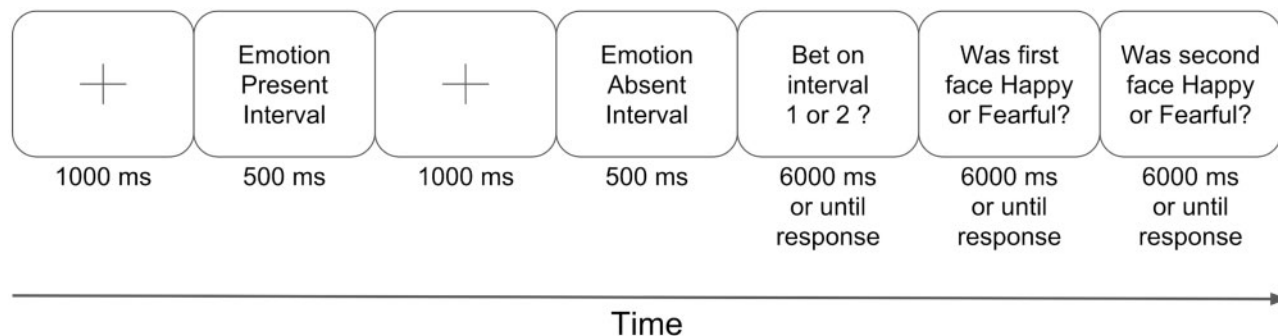

Figure 1. A sample 2-interval forced-choice (2IFC) trial. Subjects viewed two intervals containing a face with Emotion Present (EP interval) or Emotion Absent (EA interval), then "bet" on which emotion discrimination they thought they were more likely to get correct. Subjects then indicated the emotion they identified in the first and second intervals, respectively. The order of the Emotion Present (EP) and Emotion Absent (EA) intervals was counterbalanced between trials.

the subject failed to respond to either of these questions in the trial. We excluded from further analysis any subjects who retained fewer than 80% valid trials after the two exclusion criteria above were implemented.

To look for emotion processing "featural blindsight," we follow the definitions introduced by Peters and Lau (2015). If subjects show no emotion processing "featural blindsight," as soon as they are able to meaningfully identify the emotion in the EP interval ("% correct emotion discrimination" > 50%), they should exhibit some degree of introspective access evinced by an ability to bet on their choices ("% bet on EP interval" > 50%); this behavior would match predictions of a Bayesian ideal observer (Peters and Lau 2015). If, in contrast, subjects do exhibit "featural blindsight," there will be some level of objective performance capacity ("% correct emotion discrimination" > 50% in the EP interval) at which subjects have *no* introspective access to their decisional processes, i.e., *no* ability to meaningfully bet on their choices ("% bet on EP interval" = 50%). To arbitrate between these hypotheses, we compare the goodness of fit of a Bayesian ideal observer computational model to that of a Bayesian observer with additional Type 2 noise, which exhibits featural blindsight. See next sections for details.

During the calculation of d', whenever the subject exhibited a hit rate or false alarm rate of 0 or 1, we applied a standard correction (Wickens, 2001) to ensure that the calculated d' would be within a reasonable range (i.e. not infinity). For the hit rate (HR), we corrected the probability using the formulas

$$\text{If } HR = 0, \text{ then } HR = \frac{1}{nSignal + 1}$$
$$\text{If } HR = 1, \text{ then } HR = 1 - \frac{1}{nSignal + 1}$$

where *nSignal* is the number of emotion present intervals for that emotional intensity. For the false alarm rate (FAR), we corrected the probability using the formulas

$$\text{If } FAR = 0, \text{ then } FAR = \frac{1}{nNoise + 1}$$
$$\text{If } FAR = 1, \text{ then } FAR = 1 - \frac{1}{nNoise + 1}$$

where *nNoise* is the number of emotion present intervals for that emotional intensity.

Computational model

The Bayesian ideal observer computational model has previously been described elsewhere (Peters and Lau 2015). Briefly, for every 2IFC trial, we draw two samples d_{EP} and d_{EA}, each from a bivariate Gaussian distributions S_{EP} and S_{EA} ($S \sim N(\mu, \Sigma)$) representing an Emotion-Present face and an Emotion-Absent face, respectively. The mean of the generating distribution for the EP face is $\mu_{EP} = [e_{Happy}, 0]$ or $\mu_{EP} = [0, e_{Fearful}]$, depending on the random assignment of emotional valence in the EP interval; the mean of the generating distribution for the EA face is always $\mu_{EA} = [0, 0]$. $\Sigma = [1\ 0;\ 0\ 1]$, the identity matrix, for all generating distributions.

In all cases, the true emotional intensity that generated the emotion, i.e., $e_{emotion}$, is unknown to the observer; the observer only has access to the sample that is being observed. This is why the decision about whether each sample is happy or fearful is made according to Bayes' rule, marginalizing across all possible emotional intensity levels for both Happy and Fearful emotions:

$$p(S, e | d) = \frac{p(d | S, e) p(S, e)}{p(d)}. \tag{1}$$

Confidence in the decision (happy or fearful) for each interval is defined as the posterior probability of the choice that was made, i.e.,

$$p(S | e) = \int p(S, e | d) de. \tag{2}$$

The chosen emotion for each interval, both EP and EA, is defined as the emotion *i* that maximizes this posterior probability, i.e.,

$$S_{\text{chosen}} = \text{argmax}_i\ p(S_i | d) \tag{3}$$

The model then "bets" on the interval that has the higher posterior probability of the choice, i.e., higher confidence, by comparing the posterior probabilities for the chosen emotion in the EP and EA intervals via a decision variable *D*,

$$D = \log\left(\frac{p(S_{\text{chosen,EP}} | d_{EP})}{p(S_{\text{chosen,EA}} | d_{EA})}\right). \tag{4}$$

If $D \geq 0$, the model "bets" on the EP interval; if $D < 0$, the model "bets" on the EA interval.

All simulations of this model's behavior were done via custom-written scripts in Matlab (Natuck, MA), with emotional intensity level *e* ranging from 0.01 to 1 in steps of 0.01 and 5000 samples per emotional intensity level.

Simulating emotion processing "featural blindsight"
The ideal observer described above has access to Type 2 information (confidence) with the same degree of fidelity as its access to Type 1 information (emotion decision). In contrast, an observer that exhibits blindsight would have much poorer access to Type 2 information than Type 1 information. Following previous work (Peters and Lau 2015; Maniscalco and Lau 2016), this can be modeled as the addition of increasing amounts of Type 2 noise after the Type 1 decision has been reached. To simulate the behavior of such an observer, we therefore added Gaussian noise to the definition of the Type 2 decision variable *D* after the Type 1 decision, i.e.,

$$D = \log\left(\frac{p(S_{\text{chosen,EP}} | d_{EP})}{p(S_{\text{chosen,EA}} | d_{EA})}\right) + \epsilon, \tag{5}$$

where $\epsilon \sim N(0, \sigma)$. We simulated predicted behavior for increasing degrees of "featural blindsight" at values of σ from 0.01 to 1 in steps of 0.01.

Goodness of fit
The goodness of fit of both models—the Bayesian ideal observer and the "featural blindsight" observer at various levels of Type 2 noise (Eq. 5) —was calculated as the multinomial log-likelihood (L_m) of the model with parameters ϕ given the data. This measure quantifies the relative agreement between the data collected from participants and that predicted by a model. We use the following formula:

$$L_m(\phi | \text{data}) \propto \log\left(\Pi_{ij} P_\phi(R_i | S_j)^{n_{\text{data}}(R_i | S_j)}\right), \tag{6}$$

where S_j is the type of stimulus that might be shown on a given trial, and R_i refers to the behavioral response a subject produces

on that trial. $n_{data}(R_i|S_j)$ represents the count of how many times a human observer produced response R_i after viewing stimulus S_j. $P\phi(R_i|S_j)$ represents the model's prediction of the probability a subject produced response R_i after viewing stimulus S_j, according to the model with parameters ϕ, i.e., the percentage of time the model produced this "response" to this "stimulus." Importantly, the multinomial log-likelihood quantifies the fit of the *full* distribution of probabilities of each response type given each stimulus type, not just with reference to a summary statistic.

We used this metric to evaluate the predictions of both models at all levels of Type 1 performance exhibited by each human observer. Note that for the ideal observer, there are no free parameters; for the "featural blindsight" observer with Type 2 noise, the only parameter ϕ is the magnitude of the Type 2 noise (Eq. 5).

Results

Two subjects had incomplete data and our exclusion criteria led us to exclude another two subjects from our analyses, leaving 29 subjects. Across subjects, the stronger three (out of the four) levels of emotion presented typically led to above-chance performance for each emotion intensity: 5% emotion (mean % correct = 49.92 ± SD 0.053), 15% emotion (mean % correct = 0.613 ± SD 0.091), 25% emotion (mean % correct = 0.701 ± SD 0.106), and 75% emotion (mean % correct = 0.897 ± SD 0.139). A one-way repeated measures analysis of variance (ANOVA) showed the expected increase in Type 1 performance ("% correct emotion discrimination") as a function of increasing emotional intensity [$F(2.386) = 105.223$, $P < 0.001$, Greenhouse-Geisser corrected for sphericity violation; all subsequent cases with Greenhouse-Geisser correction are marked with ^{G}]. Likewise, subjects' ability to bet on the EP interval also increased with increasing emotional intensity as expected ($F(2.112) = 35.840$, $P < 0.001^G$].

The critical analysis is whether subjects show emotion processing "featural blindsight." Following the definitions introduced by Peters and Lau (2015), if subjects show such emotion processing ability without awareness, we would expect that there would be some level of objective performance capacity ("% correct emotion discrimination" > 50% in the EP interval) at which subjects had no ability to meaningfully bet on their choices ("% bet on EP interval" = 50%). To look for this possibility, we plotted the "% bet on EP interval" values against the "% correct emotion discrimination" values for each subject, and fit a smoothing spline [using Matlab's `smoothingspline` function, which minimizes the expression $p\sum_i w_i (y_i - s(x_i))^2 + (1-p)\int\left(\frac{d^2s}{dx^2}\right)^2 dx$, with the smoothing parameter set to 0.99] to the data across all subjects (Fig. 2a). Data from individual subjects closely resembles the group data (Supplementary Fig. S1), and analogous analyses using d' in the EP interval (d'EP) and d' in selecting the interval containing the emotional face (d'2IFC) revealed similar results (Supplementary Fig. S3).

From visual inspection alone, we see no evidence of emotion processing "featural blindsight": as soon as subjects are able to correctly identify the emotion in the EP interval, they are able to meaningfully bet on their choices to some degree. The pattern of the data closely mimics the failure to observe blindsight-like behavior in normal observers using low-level visual stimuli, in which the target-absent interval (here the EA interval) contained no stimulus at all (Peters and Lau 2015). It also visually

matches predicted behavior from a Bayesian ideal observer (Fig. 2b).

However, because this study was done online through Amazon mTurk, the data for individual subjects is somewhat noisier than it would be had subjects each completed many hours of psychophysics, as they did in the original Peters and Lau (2015) study (Supplementary Fig. 1). Therefore, we rely on the quantitative goodness of fit metrics to critically evaluate whether any hint of "featural blindsight" might be present.

We calculated the goodness of fit as L_m for each human observer for the Bayesian ideal observer, and compared it to L_m for the "featural blindsight" observer model at each level of Type 2 noise (see Materials and Methods, Eq. 6). This analysis revealed that as Type 2 noise increases—i.e., as "featural blindsight" becomes stronger—the model fits the data less and less well [main effect of Type 2 noise magnitude, $F(1.141) = 15.541$, $P < 0.001^G$; Fig. 2c]. This interpretation was confirmed with a two-tailed paired samples t-test of the log likelihoods between $\sigma = 0$ (no Type 2 noise, i.e., Bayesian ideal observer) and $\sigma = 1$ (high Type 2 noise) confirms that goodness of fit decreases as Type 2 noise increases [$t(28) = 3.926$, $P \le 0.001$]. Thus, the Bayesian ideal observer with no Type 2 noise—i.e. the observer with no "featural blindsight"—provides the best explanation of the human subjects' behavior.

Discussion

Here, we used a bias-free two-interval forced-choice (2IFC) method to examine whether facial emotion processing might reveal blindsight-like behavior in normal human observers. We modified a previous version of this paradigm (Peters and Lau 2015) to match low-level visual properties such as luminance, contrast, and spatial frequency differences between the Emotion Present (EP) and Emotion Absent (EA) intervals; our goal was to reveal emotion processing "featural blindsight," in which observers would be able to correctly identify an emotion when it was present but have no ability to introspect on the evidence or process leading to their choices. However, despite previous reports suggesting that emotion processing may bypass conscious visual processing areas (Pessoa and Adolphs 2010; Watanabe and Haruno 2015; Diano *et al.* 2016; Khalid and Ansorge 2017) and that emotional (especially fearful) stimuli may be powerful candidates for inducing blindsight-like behavior (Vieira *et al.* 2017), here we saw no evidence to suggest that emotion-processing "featural blindsight" may occur.

These results are in line with the absence of evidence for blindsight-like behavior reported in the original Peters and Lau (2015) study (which used forward–backward masking and simple visual stimuli), as well as the observation that various masking techniques fail to produce differences in the relationship between objective versus subjective thresholds (Knotts *et al.*, 2018). Our observations also support the finding that even non-invasive transcranial magnetic stimulation fails to produce the ability to discriminate a stimulus in the absence of any visual awareness or confidence (Peters *et al.* 2017). Taken together, the evidence appears to be mounting that it is extremely difficult—if not impossible—to use visual masking or noninvasive techniques to produce dissociations between performance and awareness that would be ideal for experimentally isolating subjective consciousness.

However, despite the consistency of the present results with previous reports, it should certainly be noted that the data collection method we used here was quite different from the standard approach. Each of our subjects completed only 112 trials in
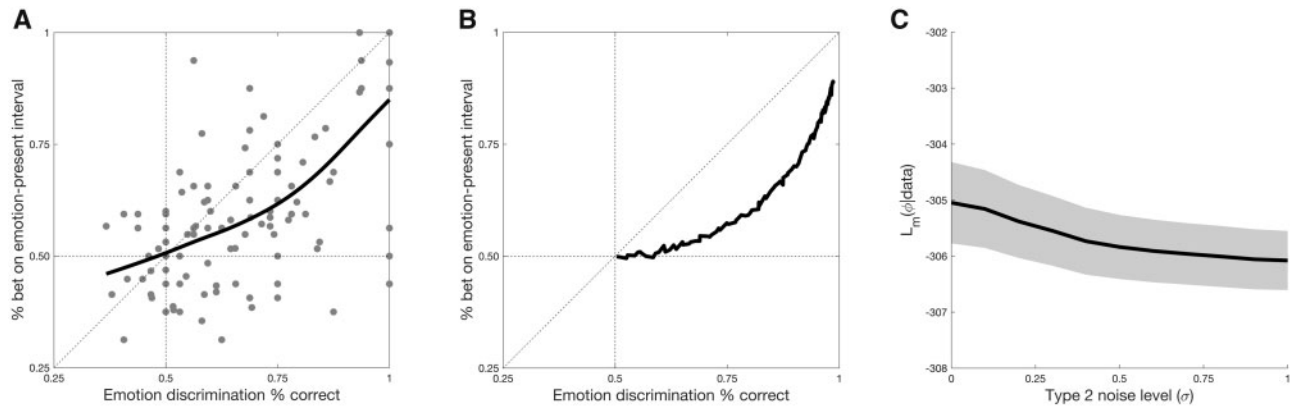
Figure 2. Human subjects show no emotion processing "featural blindsight." (**a**) As soon as subjects are able to correctly identify the emotion of the face in the Emotion Present (EP) interval ("% correct emotion discrimination" > 50%), they appear to be able to meaningfully bet on their choices ("% bet on EP interval" > 50%). Each point represents one level of emotion for one observer, such that each observer contributes four points to the plot. (**b**) Subjects' data closely matches predictions from the Bayesian ideal observer computational model. (**c**) Increasing levels of Type 2 noise (σ; see Materials and Methods) to produce emotion processing "featural blindsight" in the computational model leads to increasingly *worse* goodness of fit (L_m) between the model and the human subjects' data [$F(1.141) = 15.541$, $P < 0.001^G$; two-tailed paired samples t-test between extremes of $\sigma = 0$ [no Type 2 noise] and $\sigma = 1$ [large Type 2 noise]: $t(28) = 3.926$, $P \leq 0.001$]. Error cloud represents the standard error of the mean.

the main experiment across four levels of emotion intensity, which is significantly less than the volume of data collected in e.g. the original study (2600 trials; Peters and Lau 2015). One could argue that the online data collection method, with so few trials per subject, may preclude the possibility of detecting the blindsight-like effect we hoped to reveal. In response, we note that each of our 29 subjects' data looks much like the data in aggregate, and that the goodness of fit metrics were calculated for each subject individually rather than for the group average. That we identified decreasing goodness of fit for increasing Type 2 noise, despite the noisiness of online data collection and the small trial numbers, suggests that the data collection method was not so noisy as to prevent the possibility of identifying emotion processing "featural blindsight." Future studies may wish to follow-up the results presented here with additional variations, done both in the laboratory and in an online setting.

Another way to check whether our online data collection method produces meaningful data would be to compare the emotional discrimination results to previous reports. In another study of emotion processing and confidence, it was reported that fear processing is "special": when perceiving a fearful face human subjects show a liberal bias in both detection and discrimination tasks, and "fear" choices are also reported with higher confidence (Koizumi *et al.* 2016). We used signal detection theoretic approaches (Green and Swets 1966; Macmillan and Creelman 2004) to examine whether the same biases would be present in our own data. Consistent with previous reports that fearful faces are perceived more easily or quickly (Milders *et al.* 2006; Phelps *et al.* 2006; Stein *et al.* 2009, 2010, 2014; Stienen and de Gelder 2011; Amting *et al.* 2010), we also saw a significant shift in the Type 1 decisional criterion that biased subjects to respond "fearful" more than "happy" at lower emotional intensities in both the EP [$t_{5\%}(28) = 4.174$, $P < 0.001$; $t_{15\%}(28) = 3.917$, $P < 0.001$; $t_{15\%}(28) = 2.040$, $P = 0.051$; 75% intensity N.S.; 5% and 15% intensities are significant with Bonferroni correction for multiple comparisons] and EA [$t(28) = 5.219$, $P < 0.001$] intervals. Interestingly, the magnitude of this shift when significant or trending ($c_{EA} = -0.600$, $c_{5\%} = -0.444$, $c_{15\%} = -0.352$, $c_{15\%} =$

-0.187) was somewhat larger than that reported by Koizumi *et al.* (2016) in their discrimination task ($c \approx -0.12$). This result provides further evidence that online data collection, despite being impoverished relative to the controlled setting of the laboratory, may be a viable method for identifying even small effects.

It is interesting to note that the Bayesian ideal observer here, as in Peters and Lau (2015), may appear to show some level of performance without awareness—here in the form of featural blindsight as well—at low levels of performance, just above chance (Fig. 2b). However, we reiterate that this apparent performance without awareness is really due to a lack of statistical significance in the ability to "bet" on its choices as being different from chance in this very low region of the performance curve. In the present manuscript we implemented the Bayesian ideal observer with Monte Carlo simulations with a finite number of simulated trials for convenience, but mathematically, $p(S_{chosen}|d) > 0.5$ means that a Bayesian Ideal Observer with infinite compute – which could represent the entire posterior probability distribution—can "bet" above chance exactly as soon as it can discriminate a stimulus' identity above chance. The appearance of performance without awareness for the ideal observer therefore only comes from limitations of the simulation.

Given the nature of the 2IFC task used here in the subjective rather than objective task (de Gardelle and Mamassian 2014; Peters and Lau 2015; de Gardelle *et al.* 2016; Mamassian 2020), it is important to clearly discuss how interval betting and emotion discrimination relate to constructs of subjective versus objective processing and the impact of these relationships on the interpretation of results. Here, interval betting is used to index subjective awareness because it relates to metacognition and 2IFC detection (Mamassian 2020) [although it should not be taken to mean that we equate awareness and subjective confidence (Rosenthal 2019); in contrast, emotion discrimination capacity provides an index of perceptual performance. We acknowledge that the relationships between these measures and the constructs they assume to quantify in the present project are not necessarily universally agreed-upon, especially when it comes to the relationship between metacognition and

awareness (Nelson 1996; Rosenthal 2019). Yet these concepts have become increasingly tightly coupled both empirically and theoretically (Brown *et al.* 2019; Lau 2019; Lau and Brown, 2019), so despite their theoretical distinction it can be argued that metacognitive judgments such as those used here provide a sufficient indicator for conscious awareness in experimental contexts.

The present results—and previous studies showing similar failure to induce blindsight-like behavior using generally accepted methods for doing so (Peters and Lau 2015; Peters *et al.* 2017a)—present a growing challenge to the dominant view that stimulus manipulation and/or noninvasive brain stimulation may result in task performance capacity in the absence of introspective access (Merikle 1982; Reingold and Merikle 1988; Kolb and Braun 1995; Merikle *et al.* 2001; Boyer *et al.* 2005; Charles *et al.* 2013). We acknowledge that the growing body of work we add to here does not speak definitively to the *impossibility* of inducing dissociations between the objective and subjective thresholds in normal human observers—i.e., they do not support the possibility of inducing $d' > 0$ (above objective threshold) without accompanying awareness (below subjective threshold). However, these results do increasingly suggest that inducing blindsight-like perceptual capacity in the absence of introspective access may be much more difficult to induce than commonly believed (Heeks and Azzopardi 2015). Consequently, demonstrating such a dissociation using this conservative 2IFC confidence method would provide indisputable evidence that blindsight *can* be induced in normal observers if one employs the appropriate experimental manipulations.

Supplementary data

Supplementary data is available at NCONSC Journal online.

Data accessibility statement

The code used in the simulations and data presented here can be found at https://github.com/vrsivananda/Faces_2IFC_Task.

Conflict of interest statement. None declared.

References

Amting JM, Greening SG, Mitchell DGV. Multiple mechanisms of consciousness: the neural correlates of emotional awareness. *J Neurosci* 2010;**30**:10039–47.

Aru J, Bachmann T, Singer W, *et al*. Distilling the neural correlates of consciousness. *Neurosci Biobehav Rev* 2012;**36**:737–46.

Azzopardi P, Cowey A. Is blindsight like normal, near-threshold vision? *Proc Natl Acad Sci USA* 1997;**94**:14190–4.

Azzopardi P, Cowey A. Blindsight and visual awareness. *Conscious Cogn* 1998;**7**:292–311.

Balsdon T, Azzopardi P. Absolute and relative blindsight. *Conscious Cogn* 2015;**32**:79–91.

Block N. Two neural correlates of consciousness. *Trends Cogn Sci* 2005;946-52.

Boyer JL, Harrison S, Ro T. Unconscious processing of orientation and color without primary visual cortex. *Proc Natl Acad Sci USA* 2005;**102**:16875–9.

Brown R, Lau H, LeDoux JE. Understanding the higher-order approach to consciousness. *Trends Cogn Sci* 2019;**23**:754–68.

Charles L, King J-R, Dehaene S. Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J Neurosci* 2014;**34**:1158–70.

Charles L, Van Opstal F, Marti S, *et al*. Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage* 2013;**73**:80–94.

de Gardelle V, Le Corre F, Mamassian P. Confidence as a common currency between vision and audition. *PLoS One* 2016;**11**: e0147901.

de Gardelle V, Mamassian P. Does confidence use a common currency across two visual tasks? *Psychol Sci* 2014;**25**:1286–8.

Diano M, Celeghin A, Bagnis A, *et al*. Amygdala response to emotional stimuli without awareness: facts and interpretations. *Front Psychol* 2016;**7**:2029.

Eriksen CW. Discrimination and learning without awareness: a methodological survey and evaluation. *Psychol Rev* 1960;**67**: 279–300.

Evans S, Azzopardi P. Evaluation of a "bias-free" measure of awareness. *Spatial Vis* 2007;**20**:61–77.

Galvin SJ, Podd JV, Drga V, *et al*. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon Bull Rev* 2003;**10**:843–76.

Giles N, Lau H, Odegaard B. What type of awareness does binocular rivalry assess? *Trends Cogn Sci* 2016;**20**:719–20.

Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, Inc., 1966. https://psycnet.apa.org/record/1967-02286-000.

Hannula DE, Simons DJ, Cohen NJ. Imaging implicit perception: promise and pitfalls. *Nat Rev Neurosci* 2005;**6**:247–55.

Heeks F, Azzopardi P. Thresholds for detection and awareness of masked facial stimuli. *Conscious Cogn* 2015;**32**:68–78.

Jachs B, Blanco MJ, Grantham-Hill S, *et al*. On the independence of visual awareness and metacognition: a signal detection theoretic analysis. *J Exp Psychol* 2015;**41**:269–76.

Khalid S, Ansorge U. Subliminal face emotion processing: a comparison of fearful and disgusted faces. *Front Psychol* 2017;**8**: 1028.

Knotts JD, Lau H, Peters MAK. Continuous flash suppression and monocular pattern masking impact subjective awareness similarly. Atten Percep Psychophys 2018:80:1974–1987.

Koizumi A, Mobbs D, Lau H. Is fear perception special? Evidence at the level of decision-making and subjective confidence. *Soc Cogn Affect Neurosci* 2016;**11**:1772–82.

Kolb FC, Braun J. Blindsight in normal observers. *Nature* 1995; **377**:336–8.

Kunimoto C, Miller J, Pashler H. Confidence and accuracy of near-threshold discrimination responses. *Conscious Cogn* 2001; **10**:294–340.

Lau H. Consciousness, metacognition, & perceptual reality monitoring. PsyArxiv 2019; doi: 10.31234/osf.io/ckbyf.

Lau H, Brown R. The emperor's new phenomenology? The empirical case for conscious experiences without first-order representations. *Blockheads!: Essays on Ned Block's Philosophy*. 2019. https://books.google.ca/books? hl=en&lr=&id=bxuFDwAAQBAJ&oi=fnd&pg=PA171&ots=qMINjBZ8m4&sig=mjeqPGEMOD3k_YzR_Y_HwARyw28.

Lau HC. Are we studying consciousness yet? In: Weiskrantz L and Davies M (eds.), *Frontiers of Consciousness*. Oxford University Press, 2008, 2008–245. https://books.google.com/books?hl=en&lr=&id=okSnxl4uvYMC&oi=fnd&pg=PA245&dq=lau+2008+consciousness&ots=GPtdlk6i87&sig=hrsL49bKEJQcPYVogNrra7jXbq0#v=onepage&q=lau%202008%20consciousness&f=false.

Lau H, Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci USA* 2006;**103**:18763–8.

Lloyd DA, Abrahamyan A, Harris JA. Brain-stimulation induced blindsight: unconscious vision or response bias? *PLoS One* 2013;**8**:1–16.

Macmillan NA, Creelman CD. *Detection Theory: A User's Guide.* Taylor & Francis, 2004. https://psycnet.apa.org/record/2004-19022-000.

Mamassian P. Confidence forced-choice and other metaperceptual tasks. *Perception* 2020;**49**:616–35.

Maniscalco B, Lau H. The signal processing architecture underlying subjective reports of sensory awareness. Neurosci Conscious 2016;**1**:1

Merikle PM. Unconscious perception revisited. *Percept Psychophys* 1982;**31**:298–301.

Merikle PM, Smilek D, Eastwood JD. Perception without awareness: perspectives from cognitive psychology. *Cognition* 2001;**79**:115–34.

Milders M, Sahraie A, Logan S, *et al.* Awareness of faces is modulated by their emotional meaning. *Emotion* 2006;**6**:10–7.

Morales J, Chiang J, Lau H. Controlling for performance capacity confounds in neuroimaging studies of conscious awareness. *Neurosci Conscious* 2015;**1**:niv008.

Morgan MJ, Mason AJS, Solomon JA. Blindsight in normal subjects? *Nature* 1997;**379**:401–2.

Nelson TO. Consciousness and metacognition. *Am Psychol* 1996; **51**:1–15.

Pessoa L, Adolphs R. Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. *Nat Rev* 2010;**11**:773–83.

Peters MAK, Fesi J, Amendi N, et al. Transcranial magnetic stimulation to visual cortex induces suboptimal introspection. *Cortex* 2017a;93:119–132

Peters MAK, Kentridge RW, Phillips I, *et al.* Does unconscious perception really exist? Continuing the ASSC20 Debate. *Neurosci Conscious* 2017b;**1**:nix015 doi: 10.1093/nc/nix015.

Mak, Lau H P. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. eLife. 4:e09651 2015;doi: 10.7554/eLife.09651

Peters MAK, Ro T, Lau H. Who's afraid of response bias? *Neurosci Conscious* 2016;**1**:niw001.

Phelps EA, Ling S, Carrasco M. Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol Sci* 2006;**17**:292–9.

Ramsøy TZ, Overgaard M. Introspection and subliminal perception. *Phenomenol Cogn Sci* 2004;**3**:1–23.

Rees G, Kreiman G, Koch C. Neural correlates of consciousness in humans. Nat Rev Neurosci 2002;**3**:261.

Reingold EM, Merikle PM. Using direct and indirect measures to study perception without awareness. *Percept Psychophys* 1988; **44**:563–75.

Robichaud L, Stelmach LB. Inducing blindsight in normal observers. *Psychon Bull Rev* 2003;**10**:206–9.

Rosenthal D. Consciousness and confidence. *Neuropsychologia* 2019;**128**:255–65.

Snodgrass M, Shevrin H. Unconscious inhibition and facilitation at the objective detection threshold: replicable and qualitatively different unconscious perceptual effects. *Cognition* 2006; **101**:43–79.

Stein T, Peelen MV, Funk J, *et al.* The fearful-face advantage is modulated by task demands: evidence from the attentional blink. *Emotion* 2010;**10**:136–40.

Stein T, Seymour K, Hebart MN, *et al.* Rapid fear detection relies on high spatial frequencies. *Psychol Sci* 2014;**25**:566–74.

Stein T, Zwickel J, Ritter J, *et al.* The effect of fearful faces on the attentional blink is task dependent. *Psychon Bull Rev* 2009;**16**: 104–9.

Stienen BMC, de Gelder B. Fear modulates visual awareness similarly for facial and bodily expressions. *Front Hum Neurosci* 2011;**5**:132.

Tononi G, Koch C. The neural correlates of consciousness. Ann NY Acad Sci 2008;**1124**:239.

Tottenham N, Tanaka JW, Leon AC, *et al.* The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res* 2009;**168**:242–9.

Vieira JB, Wen S, Oliver LD, *et al.* Enhanced conscious processing and blindsight-like detection of fear-conditioned stimuli under continuous flash suppression. *Exp Brain Res* 2017;**235**:3333–44.

Watanabe N, Haruno M. Effects of subconscious and conscious emotions on human cue-reward association learning. *Sci Rep* 2015;**5**:8478.

Weiskrantz L. *Blindsight: A Case Study and Implications.* Oxford University Press, 1986. https://journals.sagepub.com/doi/pdf/10.1177/00030651970450031209.

Weiskrantz L. Blindsight revisited. *Curr Opin Neurobiol* 1996;**6**: 215–20.

Wickens TD. *Elementary Signal Detection Theory*, 1st edn. Oxford University Press, 2001. https://books.google.com/books?hl=en&lr=&id=sQ1nDAAAQBAJ&oi=fnd&pg=PR11&dq=Elementary+Signal+Detection+Theory+Wickens&ots=84lXKhVbpE&sig=Z-tfTakN4qSpe1JI-dECuX99mBM#v=onepage&q=Elementary%20Signal%20Detection%20Theory%20Wickens&f=false.

Willenbockel V, Sadr J, Fiset D, *et al.* Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods* 2010;**42**: 671–84.