Current Research in Behavioral Sciences 4 (2023) 100109

Contents lists available at ScienceDirect

Current Research in Behavioral Sciences

journal homepage: www.elsevier.com/locate/crbeha

Unbiased individual unconsciousness: Rationale, replication and developing applications

Myron Tsikandilakis[a,b,c,*], Persefoni Bali[a], Alexander Karlis[d], Pierre-Alexis Mével[c], Christopher Madan[a], Jan Derrfuss[a], Alison Milbank[e]

[a] School of Psychology, University of Nottingham, UK
[b] Medical School, Faculty of Medicine and Health Sciences, University of Nottingham, UK
[c] School of Cultures, Languages and Area Studies, University of Nottingham, UK
[d] Department of Physics, National and Kapodistrian University of Athens, Greece
[e] Department of Theology and Religious Studies, University of Nottingham, UK

ARTICLE INFO

Key words:
Unbiased
Individual
Subjective
Unconscious
Research
Methodology

ABSTRACT

Unbiased individual unconsciousness is a methodology that involves non-parametric receiver operating characteristics and Bayesian analyses and can enable a researcher to define subjective thresholds for visual suppression. It can enable a researcher to define among brief durations (e.g., 8.33 or 16.67 or 25 ms), per participant and elicitor type, the threshold of presentation for which each participant is individually unconscious during masking. The outcomes of this method are then used in a subsequent experimental session that involves psychophysiological assessments and participant ratings to explore evidence for unconscious processing and emotional responsivity. Following collegial requests for a dedicated manuscript on the rationale and replication of this method, in this manuscript, we provide a thorough, comprehensive and reader-friendly manual for this methodology. We include empirical illustrations, open-source and ready-to-use methodological, mathematical and statistical coding scripts and step-by-step instructions for replicating key parts or the entire method. We discuss the potential contributions and the developing applications of individual metrics for unconsciousness.

Introduction

Empirical research on the unconscious is an area in which researchers have provided an abundance of fascinating results and impassioned refutations. It is a research area involving intense, unrelenting and unresolved academic debates. Research in the unconscious was contentious since its "first steps" (Ebbinghaus, 1908; Field et al., 1922; Miller, 1942; Kahn, 1943; Fechner, 1948; Goldiamond, 1958; for an overview see Tsikandilakis et al., 2019b) to its – so to speak – "mid-life crisis" (Burnham, 1967; Dixon, 1971; 1981; Goodkin and Phillips, 1980; Merikle and Cheesman, 1987; Frosh, 1989; for an overview see Tsikandilakis et al., 2021a) and has grown methodologically contentious, now more than ever, among contemporary psychologists (see Bar and Biederman, 1998; Erdelyi, 2004; Pessoa and Adolphs, 2010; Elgendi et al., 2018; for an overview see Tsikandilakis et al., 2022c).

The reasons for this topical discontent have been attributed to how different our empirical outcomes for unconscious processing are and how polemically the believers and disbelievers of these outcomes hold on to their theses and antitheses (Pessoa and Adolphs, 2010; Stafford, 2014). Several scholars explored as best they could the scientific causes of this topical discontent (Stanislaw and Todorov, 1999; Erdelyi, 2004; Dienes, 2014; 2015; 2016). They identified potential problems and attempted to provide possible resolutions (see van der Ploeg et al., 2017).

The problems that scholars recognized as regards the empirical exploration of the unconscious were numerous. These included several issues, such as the use of biased metrics for the assessment of perception during visual suppression (Stanislaw and Todorov, 1999; Zhang and Mueller, 2005; Swets, 2014; Hautus et al., 2021) and the use of inconclusive statistical procedures for inferring whether participants were unconscious of visually suppressed stimuli (Dienes, 2014; 2015; 2016; Kruschke and Liddell, 2018; Heck et al., 2022). These also spanned to other issues, such as the use of potentially unreliable methods for implementing psychophysics-related image processing manipulations, the type of masking applied to visually suppressed stimuli, and the unresolved problem of failing to achieve unbiased evidence for unconsciousness using static durations of presentation (e.g., 8.33 or 16.67 or 25 or 33.33 ms) (see Tsikandilakis et al., 2019b, 2020a, 2021a, 2022c). The explicit outlining of these tangible problems had a paradoxical effect (Bargh and Morsella, 2008): it restored a quantum of

* Corresponding author at: School of Psychology, School of Medical and Health Sciences, University of Nottingham.
E-mail address: Myron.tsikandilakis3@nottingham.ac.uk (M. Tsikandilakis).

https://doi.org/10.1016/j.crbeha.2023.100109
Received 6 October 2022; Received in revised form 4 April 2023; Accepted 5 April 2023

confidence in the future of research into the unconscious. It signified a passage from thesis-affiliated conflicts to an objective exploration of the methodological issues that required to be addressed and resolved in this area.

For example, a very contentious issue in this area was what metrics should be used for the assessment of perception during visual suppression (Swets, 2014). Relevant research, in its vast majority, provided hit-rate outcomes – the percentage of correct answers – to assess perception under conditions of visual suppression (Brooks et al., 2012; Meneguzzo et al., 2014). This was problematic. Hit rates can be affected by response bias. For example, some participants might respond conservatively, that is, they withhold a post-trial engagement task response for seeing an elicitor unless they are absolutely certain. Others might respond liberally, that is, they might respond in a post-trial engagement task seeing a masked elicitor even if they are quite unsure (Stanislaw and Todorov, 1999).

To address this issue, unbiased metrics for detection sensitivity such as signal detection theory receiver operating characteristic (ROC) were suggested as an alternative (Zhang and Mueller, 2005). The advantage of receiver operating characteristics was that they took into account error for the assessment of participant responses. They provided a ratio between hits, such as true positives, signifying that a participant responded in a post-trial task that a target was presented when a target was presented, and miss responses, such as false alarms, signifying that a participant responded that a target was presented when a target was not presented (Hautus et al., 2021). Receiver operating characteristics were not biased by conservative or liberal response strategies and criteria and could provide a reliable perception metric for whether the presented masked elicitors were consciously perceived or not (see Yonelinas and Parks, 2007).

Another issue was the statistical analyses used for inferring whether participants were unconscious (Dienes, 2014; 2015; 2016). The vast majority of topical research used significance testing – in this case, a one-sample *t*-test – to infer unconscious processing (Brooks et al., 2012). This analysis allowed researchers to compare the participants' responses to chance (e.g., 50% or d' = 0 or A' & A'' = 0.5 or $A = 0.5$; see Hautus et al., 2021). The researchers explored whether they were able to report non-significant differences between the participants' detection performance and chance, and if they did, they claimed that the presented elicitors were processed unconsciously (Dienes, 2015).

As our group (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019, Tsikandilakis et al., 2019a,b, 2020a,b,c, 2021a,b, 2022a,b,c) and other researchers (Howard et al., 2000; Rouder et al., 2007, 2009; Dienes, 2014, 2015, 2016; Dienes et al., 2018; Vadillo et al., 2021) have repeatedly emphasized, significance testing can only provide evidence that we can reject or fail to reject the null hypothesis. Even if we fail to reject the null hypothesis, in this instance, we do not have evidence that the participants' performance was proximate to or at-chance. We have evidence that suggests that we can reject that the participant's perception was not significantly different to chance. This is subsequently mistreated as evidence for the participant's perception being proximate to or at chance (see Kruschke, 2011).

This issue can be resolved using Bayesian analysis. Bayesian analysis requires the standard error of a sample, the mean difference of a sample, a comparison value, and lower and upper bounds that define the range for proximity of the population value to the comparison value called credible intervals (see Dienes, 2016). With these, we can provide a Bayes factor (BF). The BF can predict the likelihood that the data were observed under the alternate hypothesis or the null hypothesis. A BF < 0.33, for example, would signify that the participants' responses provided direct evidence that the data can be predicted and observed under the null hypothesis and that they were within a priori defined credible intervals for unconsciousness (Rouder et al., 2018). In the case of $0.33 \leq BF \leq 3$, the data are considered inconclusive, and at $B > 3$, the likelihood is that the data can be predicted and observed under the alternate hypothesis, and the participants were conscious of the presented elicitors (see Rouder et al., 2007; for the presented Bayesian metrics above see http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm).

These concepts have been frequently discoursed (see Shanks et al., 2021; Tsikandilakis et al., 2021) but infrequently applied (van der Ploeg et al., 2017; see particularly Tsikandilakis et al., 2022c; pp. 16–17). This shows the extent of empirical bias in relevant research. It is also challenging from an educational perspective. Receiver operating characteristics were first introduced by Marcum in 1947. They have been revised and refined multiple times (Peterson et al., 1954; Pollack and Norman, 1964; Banks, 1970; Lord, 1985; Macmillan, 1993; Stanislaw and Todorov, 1999) until – arguably (see particularly Verde et al., 2006) – Zhang and Mueller (2005) implemented the most unbiased to date non-parametric sensitivity index (A) including several corrections to d, A' and A'' (see Tsikandilakis et al., 2019; pp. 14–21).

Sensitivity index A – compared to d' – is a non-parametric sensitivity index, that does not involve any requirements or assumptions for the normality of the underlying signal-to-noise likelihood distributions (see Maniscalco and Lau, 2014; pp. 29–32). Sensitivity index A, additionally, includes diagonal Euclidean corrections for scores that lie in the upper left quadrant of the ROC space (i.e., False Alarms Rate ≤ 0.5 and Hit Rate ≥ 0.5) that are not included in the original A' algorithmic framework (Pollack and Norman, 1964) and its subsequent revisions (A''; W.D. Smith, 1995). Sensitivity index A also includes unique mathematical functions for extrapolating from a single observation – or minimal observations – ROC areas that show the possible range for sensitivity, specificity and accuracy scores for perceptual performance using Euclidean conjectures from the reported hit-rate performance to +oo, and from +oo through the reported hit-rate performance to a hits to miss ratio = 0; called areas A1 and A2 respectively (for an open-code illustration see https://osf.io/q7kep; see also Zhang and Mueller, 2005; pp. 203–207).

Conversely, the Bayesian theorem dates back 259 years (Bayes and Hume, 1763). It was first suggested as an alternative analysis to one-sample significance testing 76 years ago by C.W. Churchman (1946). It has been suggested to be a more valid alternative to one-sample t-tests by a plethora of authors in a plethora of publications thereafter (Lindley, 1957; Lann, 1959; Rozeboom, 1960; Edwards et al., 1963; Bernardo, 1980; Berger and Sellke, 1987; Howard et al., 2000; Rouder et al., 2007; 2009; see particularly Dienes, 2014; 2015; 2016; 2019; 2021). ROC and Bayesian analyses are for the better part of the last thirty years included in standard statistical software used for psychological research (SPSS, STATA, MEDCALC, etc.). We (https://osf.io/sfwbk/), and other researchers (https://osf.io/4gfwn/), have made available open-access easy-to-populate toolboxes for implementing ROCs and Bayesian analyses. This – to phrase our argument as provocatively as it should be phrased – has left a large number of topical researchers unmoved and has created a niche of relevant research in which the researchers are possibly unaware of or unwilling to take these methods under consideration. With a mind that patience is a virtue, we have reiterated these themes and resolutions here. We will continue to iterate them until common sense becomes common practice (Tsikandilakis et al., 2019b, 2020a, 2021a)

These resolutions have offered us reliable metrics for perception and appropriate statistical models for assessing unconsciousness when doing research (see Chambers, 2019; see particularly Bargh and Hassin, 2021). They have also raised new challenges. For example, using unbiased metrics and analyses for unconsciousness, attaining unconsciousness became significantly harder (see Elgendi et al., 2018). In other (and again provocative) words, it was not necessarily a difficult task when using biased metrics for perception (i.e., hit rates), and misusing statistical inference (significance testing one-sample t-tests) to suggest that stimuli were presented without conscious awareness (see Dienes, 2015). On the other hand, when using unbiased receiver operating characteristics (i.e., sensitivity index A; Zhang and Mueller, 2005) and statistical analyses that require the conclusive provision of evidence for the likelihood that the data were observed under the null (i.e., Bayesian analyses; Dienes, 2014), inferring unconsciousness became considerably more difficult (Moore, 2008).

To address this hurdle, we previously applied a method for attaining unbiased individual unconsciousness (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018; Tsikandilakis et al., 2020a). We also synoptically discoursed this method (Tsikandilakis et al., 2019b, 2021a). This method addressed the difficulties for attaining unbiased evidence for unconscious presentations. It was conceptually founded on certain empirical outcomes. It was founded on empirical outcomes that suggested that different participants required different durations of presentation for attaining unaware presentations during masking (Pessoa et al., 2002; Pessoa, 2005; Japee et al., 2009; Albrecht et al., 2010; Albrecht and Mattler, 2012). It was also founded on empirical outcomes that different elicitor types required different durations of presentation for attaining unaware presentations during masking (Pessoa et al., 2002; Pessoa and Ungerleider, 2004; Balconi and Lucchiari, 2007; Hedger et al., 2015; Elgendi et al., 2018).

The justifications of these differences in perceptual ability have been attributed to differences in attentional and cognitive resources of individual participants (Barrett et al., 2004; Bishop, 2008; Kaspar and König, 2012). They have also been attributed to the evolutionary value (Öhman, 2009; but see also Leong et al., 2022), low-spatial frequencies (De Gardelle and Kouider, 2010) and high-level component characteristics (Guenter et al., 1998) of different elicitor types that are discoursed at length in previous publications (see for example Tsikandilakis et al., 2019b, 2021a, 2022c).

The overarching aim of the current work is not merely to review these – arguably (Tsikandilakis et al., 2022c) – known and extensively discoursed subjects (see Pessoa et al., 2002; Pessoa, 2004; Japee et al., 2009; Pessoa and Adolphs, 2010); it is to show how to overcome them. We present here in full the rationale and a methodology for replication for attaining unbiased individual unconsciousness. Individual unconsciousness employs unbiased ROC metrics and Bayesian analyses, and can enable us to adjust the threshold of presentation of a masked elicitor separately for each participant for each presented elicitor type to chance-level detection performance, that would signify that a participant responded like a blind person would (see Erdelyi, 2004; pp. 76–81; see also Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019b, 2020a).

The methodological framework of this method is based on dividing a study into two stages. These two stages were separated by a time of interval of seven days. This was implemented as such because in our early pilots we showed evidence that Bayesian evidence for equivalence of significance for the null, in this case, that perceptual performance was proximate between the two sessions, was strongest (BF < 0.11) at that particular interval compared to other – shorter and longer – alternative time intervals we explored (see supplementary materials in Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018; 10.1016/j.concog.2017.10.013). In this case, we followed our own pilot findings and that does not mean that researchers who are interested in this method should not rely on their own empirical outcomes for implementing optimal between sessions time intervals.

In stage one, a set of masked stimuli is presented. These stimuli are presented for different durations, such as 8.33 and 16.67 and 25 ms. The participants are asked whether they perceived a masked target after each trial. The duration of presentation that for each participant and elicitor type provides evidence for the null when subjected to Bayesian analyses using unbiased sensitivity index A is selected as the threshold of unconsciousness for that participant and for that elicitor type (see for example Tsikandilakis et al., 2018,2020a).

In a second stage, at the same timeslot and exactly week after the first stage (see Garrido et al., 2020), the same participants are presented with different stimuli belonging to the same elicitor types (e.g., happy, fearful and neutral faces). These stimuli are presented using masking for the durations that were shown to provide substantial evidence for the null in stage one. Participants' self-reports, such as ratings for valence and intensity, and psychophysiological responses, such as skin conductance responses (SCR), heart-rate responses (HR), facial-expression assessment (FA), and further physiological and neuroscientific assessments are measured to explore unconscious processing and emotional responsivity (see Tsikandilakis et al., 2020a).

This methodological approach has received an academic welcome (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2019; Tsikandilakis et al., 2020a). It has been previously synoptically discoursed (Tsikandilakis et al., 2019b, 2021a). It would not be an understatement to submit that its reception raised considerable interest and debate concerning exactly how to replicate it, which participant inclusion criteria should be used, what were the most appropriate coding methodological and statistical scripts that should be employed, what should be the methodology for measuring and interpreting the psychophysiological outcomes of this method, and, in general, a demand for an exact, detailed and thorough discourse of this methodology (Madipakkam and Rothkirch, 2019; Maldonado et al., 2020). As requested, therefore, in this manuscript, we provide the reader with a step-by-step rationale and replication guide, including open-source statistical and methodological coding scripts, and empirical data illustrations, to enable a direct replication of key parts or of the entire methodology for unbiased individual unconsciousness.

Making stimuli unconscious

Let us assume that in stage one we would like to present participants with masked fearful, happy and neutral faces and adjust per participant and elicitor type the individual threshold for unconsciousness. The first thing we need to correctly calculate is power. For our intended design, liberal power calculations ($P_{(1-\beta)} = 0.8$; $p \leq 0.05$; $f = 0.25$) require $n = 28$ and conservative power calculations ($P_{(1-\beta)} \geq 0.9$; $p \leq 0.01$; $f = 0.1$) require $n = 302$ (Faul et al., 2009). Obviously, we would, if it is possible, advise for the latter, but a compromise between these extreme lower and higher computations ($P_{(1-\beta)} = 0.9$; $p \leq 0.01$; $f = 0.25$) requiring $n = 49$ has proven sufficient in previous studies to provide us with meaningful results (see Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018,2020a).

We need to bear in mind two more things. Firstly, we would not object to the statistical scholar who would consider the default $\varepsilon = 1$ sphericity value coded in standard power calculation software (e.g., GPOWER; see Faul et al., 2009) an "ideal world" value (see Berkovits et al., 2000). The overall mean of ε in our previous studies ($k = 17$) was 0.83 (SD = 0.04). Replacing the default coded value $\varepsilon = 1$ with $\varepsilon = 0.83$ in GPOWER 3.1 gives us $n = 58$. This increase in the sample size could contribute to an increase of the validity of our outcomes (Lane, 2016). Secondly, we should not forget the importance of contour in validating power calculations (Baker et al., 2020). In our case, for our power calculations, 40 to 60 trials per condition (i.e., elicitor type) must be presented to achieve a $P_{(1-\beta)} = 0.9$. This is a very understated and underused statistical calculation in experimental psychology and can lead to Type-I and Type-II errors (see Cohen et al., 1993; Brand et al., 2010; Murphy et al., 2014; Arnoldo and Víctor, 2015; for calculating trial contour, see https://shiny.york.ac.uk/powercontours/). It is advisable to select from the required range of trial contour a number that can divide to an integer (e.g., 40; 40/2 = 20). This can allow participants to attain unbiased perfect chance-level performance (e.g., $A = 0.5$) if this is, indeed, their true perceptual sensitivity, and avoid unfeasible-to-implement decimal repetitions for the contour of the presented stimuli (Tsikandilakis et al., 2018) .[1]

[1] We could also calculate Bayesian power for equivalence of significance to the null for the purposes of the current stage. This is a complicated process both conceptually and as regards communicating comprehensively the required code that would provide a meaningful result (Rescorla, 2015). It would require an ad-

Our next implementations relate to sampling and particularly to participant selection criteria. We have imposed and debated very strict criteria to improve the reliability and validity of our studies (Tsikandilakis et al., 2020a). Our first step is to screen participants for psychiatric conditions. The reader can choose for this purpose from a plethora of available, validated questionnaires that have been widely used in previous research (see Haynes, and Lench, 2003). They can also choose more contemporary questionnaires that include less widely applied but potentially more up-to-date psychiatric assessment items (see Rosellini and Brown, 2021). Our choice for a clinical assessment questionnaire is Sphere-12 (Hickie et al., 2001). Our choice is made on the grounds that Sphere-12 has convergent validity with assessments for both temporary mood and undiagnosed psychiatric conditions (Berryman et al., 2012; for an open-source format of Sphere-12, see https://osf.io/u8z6w/).[2]

The next questionnaire we commonly use is an assessment for Alexithymia and Alexithymia traits. These can be described as a subclinical difficulty to express, experience and recognize emotion (Taylor and Bagby, 2000). The inclusion of participants with Alexithymia, or Alexithymia traits, exposes the results of particularly emotion-assessing studies to potential biases. This is also a little recognized fact, and assessments for Alexithymia have not been consistently applied as an exclusion criterion for the participants or the participants' data in topical research. This has resulted in lack of knowledge concerning whether the participants were potential responders or non-responders of true positives (i.e., hits) for the perception of masked elicitors (see van der Ploeg et al., 2017). Our choice for resolving this issue is participant assessment and data-exclusion of above-threshold values using the OAQ-G2 (Thompson, 2007). Our choice is made on the grounds that OAQ-G2 provides reliable assessments for the diagnosis of Alexithymia and separately for high scores for Alexithymia traits, such as emotional awareness, social detachment and interpersonal functionality (Donges et al., 2014; for an open-source format of OAQ-G2, see https://osf.io/w6bmu/).

Another assessment we commonly use is the Emotional Regulation Questionnaire (ERQ; delValle et al., 2021). The ERQ is a brief 10-item assessment involving ratings from one (strongly agree) to seven (strongly disagree). It is used to assess cognitive emotional reappraisal, such as the intensity, control and ability to access and process one's emotions, and emotional suppression, such as the inclination and predilection to communicate one's emotions (Ochsner. and Gross, 2005). We use this questionnaire on the grounds that it has been shown to report high convergent validity with the sensitivity of perception and the psychophysiological outcomes of exposure to emotional images (Mauss et al., 2005). Extreme values in this questionnaire, therefore, can be an indication that both perception and physiological responses could be influenced by subclinical emotional traits (see Tsikandilakis et al., 2020a; for an open-source format of the ERQ, see https://osf.io/w827g/).

Finally, given the sensitivity of the current design to confounds, we implement an additional and strict self-developed method (see Tsikandilakis et al., 2022a). We have termed this method Bayesian sampling. We employ it when using the ERQ, and any other questionnaire we have previously used that could provide evidence for non-clinical emotion processing biases (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2020a). Our method consists of, firstly, excluding the data from participants that have scores in the ERQ – or any other relevant emotional questionnaire (Van Humbeeck et al., 2002) – that are 1.5 units below or above the median interquartile range (IQR) for Q1 and Q3 respectively. This method allows us to control for outliers without having to rely on the assumption of parametricity (see particularly Leys et al., 2013).

Subsequent to this intervention, we reverse-engineer Bayesian analyses for equivalence of significance (see particularly Zednik and Jäkel, 2014; C. 2016). We know the mean, standard deviation and standard error of normal values in the wider population for our assessment (ERQ $_{\text{emotional regulation}}$: M = 10.49; SD = 2.91; SE = 0.09; ERQ $_{\text{emotional suppression}}$: M = 21.53; SD = 3.86; SE = 0.11; see for example Sala et al., 2012). We also know the mean for each of our participants. We do not know and cannot reliably know in this instance, without falling prey to the limitations of extrapolating Monte-Carlo simulations from a single value derived from categorical responses (Garrido et al., 2016), and re-assessment participant biases (Gross et al., 2012), the measures of dispersion for each individual participant. We can, nevertheless, test whether each participant provides Bayesian evidence for equivalence to the null for canonical values for this – or any other applied – questionnaire assessment (see Tsikandilakis et al., 2019a; pp. 6 & 12; Tsikandilakis et al., 2021a, pp. 11, 17 & 19; Tsikandilakis et al., 2022a, pp. 9 & 28; Tsikandilakis et al., 2023a; pp. 8–9).

For example, let us assume that a particular participant has an ERQ score for emotional regulation of 10.51. The mean for a canonical score including cross-cultural and mixed gender effects is 10.43. The mean difference between the participant's score and the comparison value is 0.08. The standard error of the overall assessed population is 0.09. If we are liberal with our credible intervals, we can define the lower and higher bounds at - 5.87 and 5.87 (2*SD $_{\text{overall population}}$). If we want to be quite strict with our credible intervals, we can define them at - 2.91 and 2.91 or less (such −1 and 1 intervals based on end-user justified and minimum effect size of interest values; see Dienes, 2019; Dienes, 2021). For this participant, therefore, the chances that their emotional regulation score can be observed as proximate to a controlled general population value is BF = 0.03 for liberal credible criteria and BF = 0.06 for strict criteria (BF = 0.17; CIs (−1 to 1)). This means that, even under strict inclusion control conditions, we were able to report substantial evidence that the data of this participant were predicted and observed under the null and can be included in the analyses (for open-source statistical code for this method, see https://osf.io/9e5w7/). The combination of the aforementioned methods for participant inclusivity is quite thorough and exacting but it allows us to control for outliers within our own data – without relying on the assumption of parametricity – and in comparison between our data and assessment values as reported in previous research (see Gullone and Taffe, 2012).

Our next steps relate to stimuli selection. The selected stimuli should ideally be controlled for valence, intensity and emotional ambiguity. These variables should ideally provide same-elicitor type Bayesian evidence for equivalence of significance (Tsikandilakis and Chapman, 2018). In a previous publication (Tsikandilakis et al., 2018; pp. 78–91), we introduced a metric for this type of elicitor selection. We conducted a pilot study, including 52 participants (26 female; P $_{(1-\beta)}$ ≥ 0.9; p ≤ 0.01; f = 0.25), to select from a surplus of 2000 faces

dress in an at least partly mathematically populated publication (see Schönbrodt & Wagenmakers, 2018). However, we can compute, for the purposes of the current stage, that our current design has sufficient Bayesian power for providing evidence for the null given credible intervals for equivalence of significance testing for the null conservatively defined at less than a small critical f (.1) given n = 49 (P (H_{01}) = .9; BF < .33; f (.01 to .09)) and n = 58 (P (H_{01}) ≥ .9; BF < .33; f (.01 to .09)) (Kruschke, 2011). These can be converted to indicate contribution effects of .0001% to .0009% of an interdependent to a dependent variable (see Kruschke & Liddell, 2018). Calculations, instructions and code for replicating this outcome and calculating Bayesian power have been made available online and can be found at https://osf.io/2wjxa/ and https://osf.io/gzc4e/.

[2] In the interest of applied ethical standards in scientific practice, the participants should be informed of the outcomes of the clinical assessments, and the researchers should have a document ready directing them to university and local psychological support infrastructures and clinical-assessment facilities. We stand by that the responsibilities of a research should go beyond career-advancing practices. The participants were invited to a study on the premises of the correct or incorrect knowledge of not being aware that they are suffering from a psychiatric disorder. In case we have a mind for exploring the exceptions of this pre-requirement towards a publication, we would like to stress that this could constitute an invaluable contribution to this area if it was a planned objective (Hedger et al., 2016), and we would also like to stress that this could be unethical practice given our study description (see Sinclair, 2017).

(Gur et al., 2002) the most potent elicitors for fearful and happy faces, and the neutral faces that did not confer emotional responsivity. The metric we used is presented below:

$$Selection\ Criterion\ (\%\ metric) = \Bigg(\Bigg(\Bigg(\Bigg((10 - Ambiguity\ Rating) + \frac{Valence\ Rating + Intensity\ Rating}{3}\Bigg) * 50\Bigg) + \Bigg(\Bigg(\Bigg(SCR\ Maximum\ Deferral\ \{SCR\ Maximum\ Deferral\ Stimulus\ Type\}\Bigg) * 25\Bigg) + \Bigg(\Bigg(HR\ Maximum\ Deferral\ \{HR\ Maximum\ Deferral\ Stimulus\ Type\}\Bigg) * 25\Bigg)\Bigg)\Bigg)$$

In this occasion, this signified that in a separate experiment, involving all the aforementioned sampling criteria, we inquired about the valence, intensity and emotional ambiguity and assessed skin conductance and heart-rate responses for over 2000 faces to conclude to the selected sample of faces that we used in our main study. We would not object to that this choice can be burdensome for most projects, but, more critically, that it can result in including the most emotionally arousing but not necessarily the most representative and ecologically valid emotional stimuli for each elicitor type (see Schyns and Oliva, 1997).

Another method we have used is predefining meaningful values for our included elicitor types and conducting Bayesian analyses for finding out which of our stimuli meet sensible/end-user judgement appropriate a-priori requirements (see Wagenmakers et al., 2018; Dienes, 2019; 2021). For example, for our current design, including three types of faces (fearful, happy and neutral), we can either derive a priori values from previous research (Kring and Sloan, 2007; Adolph and Alpers, 2010) or define them ourselves (see for example Maxwell et al., 2015; Wiggins and Christopherson, 2019; Tsikandilakis et al., 2020a). In both cases, assessment and validation using digital facial-emotional recognition technology, such as Noldus Face Reader (see Skiendziel et al., 2019), is a potential merit for pre-selection but not necessarily a requirement for pre-selection (see Tsikandilakis et al., 2019, 2021b, 2022a, 2023).

For example, in studies like ours in which stimuli are presented for very brief durations, we would be well advised to include assessments for valence, intensity and also emotional ambiguity (Barrett, 2006). This is because in previous studies we have shown that valence and intensity can differ within a same-elicitor-type category (Tsikandilakis et al., 2018, 2019; Tsikandilakis et al., 2020a). We have also shown that stimuli categorized under the umbrella term of a prototypical emotional category can be better described as other emotions, such as faces categorized as angry providing empirical evidence for re-categorization as hostile (Tsikandilakis et al., 2020b), and stimuli categorized as sad involving characteristics relating distinguishably to melancholy, bereavement, misery and despair (Tsikandilakis et al., 2023a; see also Cowen and Keltner, 2020). These can lead to differences in perceptual sensitivity and emotional responsivity (Tsikandilakis et al., 2021a).

In an unpublished pilot conducted in 2020, we recruited a sample size of 52 participants (26 female; $M_{age} = 23.01$; $SD_{age} = 0.94$) ($P_{(1-\beta)} \geq 0.9$; $p \leq 0.05$; $f = 0.25$). We did not want strictly prototypical stimuli because these can lead to extreme emotional responsivity and lack ecological validity (Tsikandilakis et al., 2019, 2021, 2022a, 2022b, 2022c, 2023a). For ambiguity, a value of eight in a scale of one (very ambiguous) to nine (not ambiguous at all) would be suitable as an end-user justified value (see Dienes, 2019; 2021). If our results yielded unexpectedly low ratings, a value of seven could also confer some extent of approximation to low ambiguity. Values less than seven would not fit the purpose of the analyses (see Dienes, 2019). The purpose of the Bayesian analyses was to choose stimuli that were rated with the least within-elicitor-type ambiguity (see Pauker et al., 2010). For neutral faces, a value of five was considered meaningful to reliably confer null emotional characteristics for intensity and valence. For happy faces, an intensity value of seven in a scale from one (not intense at all) to nine (very intense), and a valence of seven in a scale from one (very negative valence) to nine (very positive valence) was selected as a meaningful value. Conversely, fearful faces were required to be at proximity of an intensity value of seven and a valence value of three.

In this pilot, we did not have to reverse-engineer Bayesian analysis. Given ratings for the same image from the sum of our population size ($n = 52$), we could calculate means and measures of dispersion for each image. For example, one of our images showing a happy face was rated with mean valence of 7.13 (SD = 0.67). We used a simple transformation of measures of dispersions ($SE = \frac{SD}{\sqrt{n}}$), to calculate the standard error, which yielded $SE = \frac{.67}{7.21} = 0.09$. We applied strict (- 1 to 1) lower and higher bounds, and given a mean difference of 0.13, we reported a BF = 0.32. This value qualified the particular image for inclusion. This was an arduous task that was applied individually for every image and for each individual rating for every image (i.e., valence, intensity, ambiguity). The resulting pool of images rewarded us in that every elicitor was without exception within our predefined criteria (see Dienes, 2014; 2015; 2016; 2019; 2021).

Our next steps related to stimuli processing and presentation. These topics are very important. We have provided a dedicated paper pending for publication for these, therefore, we will keep this section concise and refer the reader to our dedicated discourse (Tsikandilakis et al., 2022c). The method we most commonly used for visual suppression was backward masking (see for example Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019b, 2020a, 2021a; see also Tsikandilakis et al., 2022c). Backward masking is a widely used technique (see van der Ploeg et al., 2017). It involves the presentation of a brief target stimulus, such as, typically. but not exclusively. emotional faces (see Axelrod et al., 2015), for approximately 6.94 to 50 ms that is subsequently followed by a noise image. The aim of the noise image is to disallow conscious perception of the target stimulus (Stein et al., 2020; see also Figs. 1& 3). The canon for processing faces when using backward masking is cropping a face to standard dimensions (e.g., height: 6 cm, width: 4 cm), removing non-facial characteristics, and applying grayscale conversions to control for luminosity and color contrast (see Brooks et al., 2012). All but the latter are common and sensible techniques (but see also Kim et al., 2010); grayscale conversions can be quite problematic (De Gardelle and Kouider, 2010). The problem with grayscale conversions is that they maintain the color contrast and luminosity of the original images and transcribe it to diverse shades of gray (see Gray et al., 2013; see also Tsikandilakis et al., 2022c).

Visual stimuli can be broadly divided as having low and high-level component characteristics. The low-level component characteristics include luminosity, brightness and contrast while the high-level components involve the distribution of these characteristics that contribute to the structure of an image (Shapley,1990; see also Wyart and Tallon-Baudry, 2009; Spillmann and Werner, 2012; Izmailov et al., 2022). We can manipulate the former and average their lumens values to exactly similar values which will result in a reduction of perceptual sensitivity bias due to low-level facial features (Willenbockel et al., 2010). If we apply this method both to our masked and mask stimuli we can have "equivalence of contrast" of the mean lumens of the elicitors in our design as illustrated in Fig. 1.

We cannot manipulate the distribution of luminosity of images. We cannot manipulate high-level component characteristics in images, such as the distributions of lumen values, without deconstructing the basic facial characteristics of an expression that confer emotion (see Gray et al., 2013; for an infamous attempt to manipulate high-level component characteristics in facial expressions of emotion, see https://twitter.com/_AlecJacobson/status/1519499811816452098). Therefore, our controls in processing and presenting stimuli are limited to averaging mean lumens values within masked elicitor types and

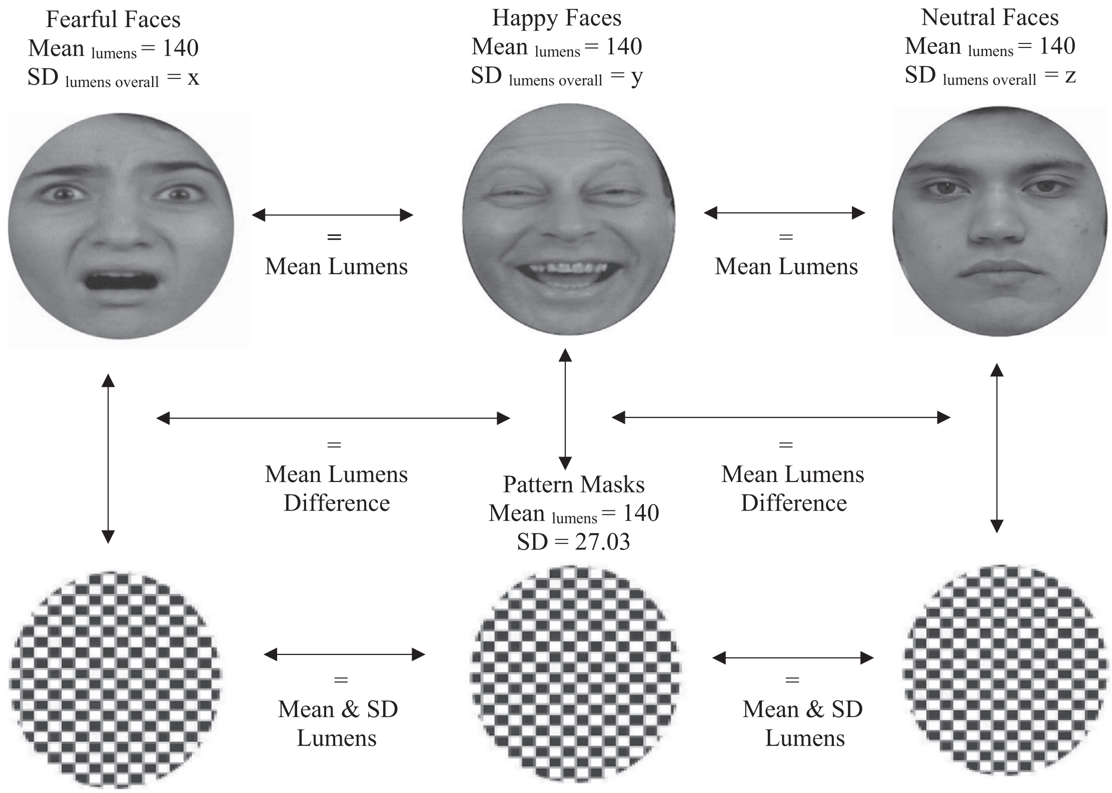

Fig. 1. Example of "Equivalence of Contrast".
Equivalence of contrast for low-level components of faces between masked and masking stimuli mean lumens using SHINE, MATLAB processing. The Mean lumens of both the pattern masks and the presented target is averaged to a mid-lumens value of 140. This can reduce variations in masked stimuli perceptibility due to differences in lumens values and masked-to-mask stimuli contrast. This figure is adapted from Tsikandilakis et al., 28).

masked-to-mask stimuli contrast as illustrated in Fig. 1 (De Gardelle and Kouider, 2010). This is not an evitable limitation and the already applied corrections have been sufficient to eliminate perceptual biases in past studies (Tsikandilakis et al., 2022c; pp. 19–33; for open-source coding scripts for averaging lumens values, see https://osf.io/xuvtd/ & https://osf.io/5zp3r/).

As we stated before, to achieve the necessary power contour for our design, each elicitor should be presented at least forty times, or in forty trials (Baker et al., 2020). Therefore, during stage one, fearful, happy and neutral faces will have to be presented forty times for each duration that in previous research has provided evidence for qualifying as a potential duration for being an individual threshold for unconsciousness (e.g., 8.33 or 16.67 or 25 ms; Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019b, 2021a). This amounts to 120 faces per elicitor type or 360 faces overall. To provide unbiased sensitivity index A outcomes we can benefit from an equal signal to noise ratio, therefore, we must also include 360 Gaussian blurs (120 blurs for 8.33 and 16.67 and 25 ms; see Tsikandilakis et al., 2022c; pp. 11–13). These will be used as the metric for our false alarms such as responding that a face was presented when a face was not presented during a binary post-trial signal-detection engagement task (i.e., "Did you see a face during the presentation? (Y/N)"; see Zhang and Mueller, 2005).

This implementation will allow us to compute the overall A for stage one. We are interested in which duration (8.33 or 16.67 or 25 ms) is per participant and elicitor type closest to chance. One way to achieve an unbiased per participant and elicitor type sensitivity A with equal signal (faces) to noise (blurs) ratio is pre-experimentally assigning a random set of forty Gaussian blurs that have the same duration as a presented elicitor type (8.33 or 16.67 or 25 ms) to that elicitor type for each duration (i.e., Fear_Blurs-Set_8.33, Happy_Blurs-Set_8.33, Neutral_Blurs-Set_8.33; Fear_Blurs-Set_16.67, Happy_Blurs-Set_16.67, Neutral_Blurs-Set_16.67; Fear_Blurs-Set_25, Happy_Blurs-Set_25, Neutral_Blurs-Set_25).

We can also do this task post-experimentally, as long as we do not match the best fit false alarm trials with an elicitor type duration that will provide optimized results for A, or fall prey to simple and innocuous human error, such as repeating one or more Gaussian blurs as part of the corresponding set of two or more elicitor types. We could also simply divide the overall number of trials (k = 120) of Gaussian blurs for each duration (8.33 or 16.67 or 25 ms) by three and include the resulting scores for false alarms to the corresponding duration of each of our elicitor types. Given the basic framework of the central limit theorem this would, in fact, work in favor of achieving proximity to $A = 0.5$. As such, it will represent a dissemblance and not a realistic equivalence of signal to noise ratio (see Zhang and Mueller, 2005; pp. 208–211). The first proposed method involving randomised pre-experimental denominations of the noise to signal ratio is the one we use in our work (see Tsikandilakis et al., 2020a).

For example, let us present the above as an empirical illustration. In our aforementioned 2020 pilot, one example – and exemplary for illustration purposes – participant was presented with forty fearful, happy and neutral faces for 8.33 and 16.67 and 25 ms with backward masking to a black-and-white pattern (Fig. 1). They were also presented with 120 Gaussian blurs for 8.33 and 16.67 and 25 ms using a pre-experimental assignment of forty blurs to each duration by elicitor type combination, as described in the first part of the previous paragraph. In Fig. 2, we can see their performance overall and for fearful, happy and neutral faces.

In this case, it must be noted, that we chose to implement an all-inclusive design, such as presenting mixed stimulus types in each stage in each session, instead of a blocked design, such as showing separately each stimulus type and an equal number of blurs, to conduct our ex-

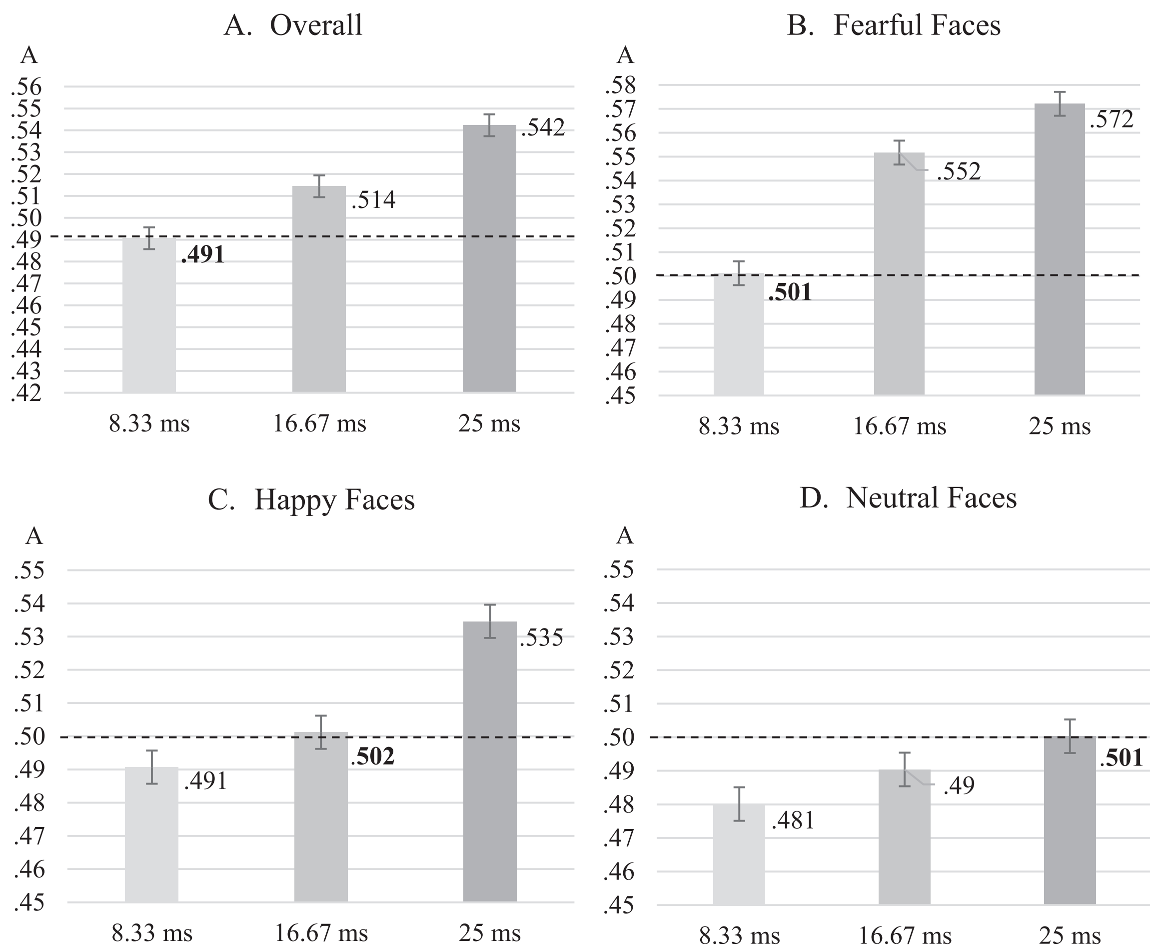

Fig. 2. ROC Performance.
Example overall ROC performance (A.) and ROC performance in response to fearful (B.), happy (C.) and neutral faces (D.) for a single participant from a previous pilot study. Dashed mid-line indicates chance-level performance at $A = 0.5$. Bars represent ±2 standard errors of the mean. Values with **bold** characters for fearful, happy and neutral faces indicate the closest to chance available sensitivity index A value from the range of included durations of presentation (8.33 and 16.67 and 25 ms). Overall perceptual performance provided Bayesian evidence for chance-level responses at 8.33 ms, nevertheless, for 8.33 ms were used for fearful faces, 16.67 ms were used for happy faces and 25 ms were used for neutral faces because these durations provided per specific stimulus type Bayesian evidence for chance-level perceptual responses. Fig. 2A-B., including overall perceptual responses and responses for fearful faces, involve wider range y axis values compared to 2C-D for illustration purposes.

periment. We do not oppose blocked designs, nevertheless, our rationale for implementing an all-inclusive design was that during false positives in all-inclusive designs only, when participants experienced physiological arousal we showed that their discrimination responses showed that they reported seeing fearful faces allowing us to previously argue for the primacy of a fear-response module during emotional misperception (see Tsikandilakis et al., 2020a). Also all-inclusive designs are suggested to require a higher cognitive load that more closely resembles circumstances of ecological validity, such as these that a person is more likely to encounter in a real-life environment when they are briefly or for-a-brief-glance exposed to multiple types of facial expressions (see Li and Wright, 2000; Sparfeldt et al., 2006; Yaremko et al., al.,2013; Kratochwill and Levin, 2014). Again, if the interested researcher would prefer blocked to all-inclusive designs, the above should stand only as a proposition on our part on how our experimental outcomes suggest that this method could be better conducted and not a methodological tenet.

For the specific participant, overall performance was closest to chance at 8.33 ms. With a mean difference to absolute chance ($A = 0.5$) at 0.009, a standard error of 0.01 and lower and upper bounds set at - 0.5 ($A = 0.45$) and 0.5 ($A = 0.55$), we reported a BF = 0.04 (for open-source software for replicating this analysis, see https://osf.io/gjaxd). This is an indication that this participant was overall proximate to chance and their performance could provide further evidence for per elicitor type unconsciousness. For fearful faces their mean difference to chance was 0.001 at 8.33 ms, for happy faces it was 0.002 at 16.67 ms and for neutral faces it was 0.001 for 25 ms. Using the exact same parameters for analyses as above, fearful faces should be presented for 8.33 ms (BF = 0.03), happy faces presented for 16.67 ms (BF = 0.03) and neutral faces presented for 25 ms (BF = 0.03).

It is worth noting that neutral faces involved the highest threshold for chance-level perception (25 ms) and the lowest means for perceptual performance; being for most durations below chance-level perception (see Fig. 2D: Neutral Faces). This is not and should not be interpreted simply as noise (see Hess et al., 2016). It is a finding that we have reported before (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019b, 2020a). It is a finding that other research groups have reported before (Breitmeyer and Ogmen, 2000; Carlson et al., 2011; Pegna et al., 2011; Rassovsky et al., 2011; Besken and Mulligan, 2013).

Previous reviews have attributed this effect to masking neutral faces with neutral faces, and thus creating masked to mask stimuli match and mismatch variations in perceptual performance; which could not be the case in the current illustration due to the use of pattern masks

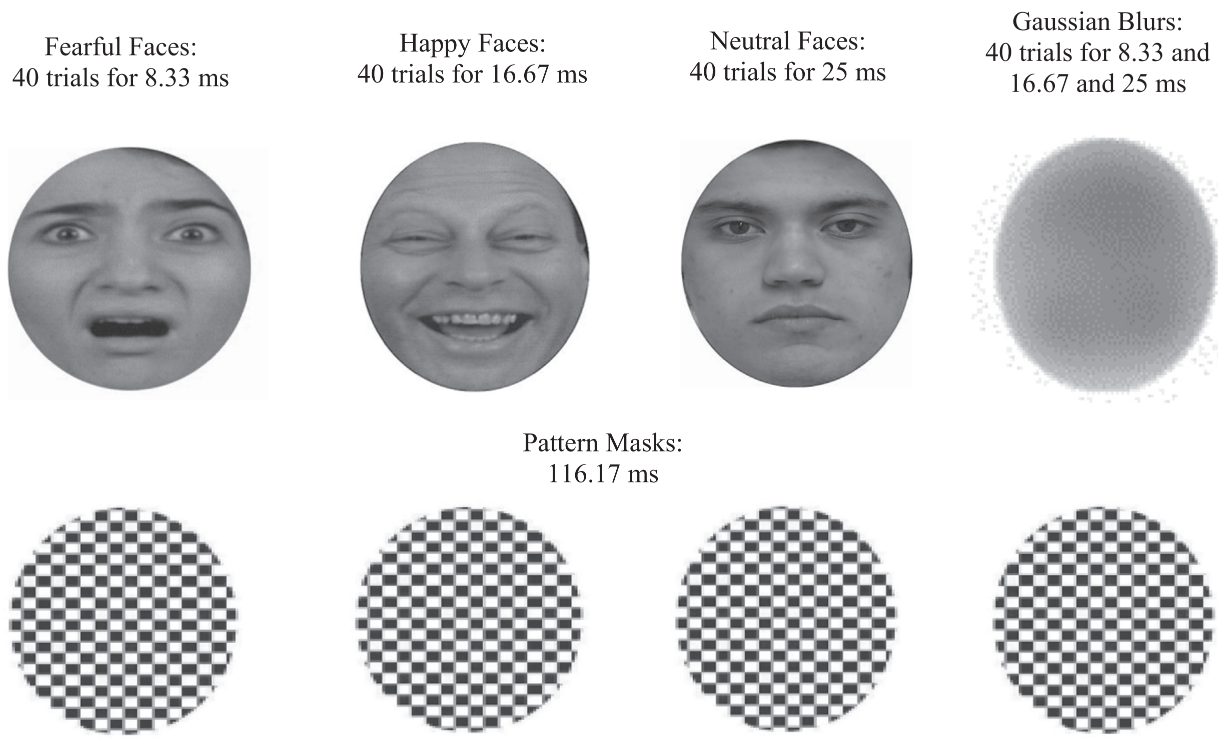

Fig. 3. Experimental Design for Stage Two.
Experimental design for stage two. Forty fearful faces for 8.33 ms, forty happy faces for 16.67 ms and forty neutral faces for 25 ms. 120 Gaussian blurs were also presented matched to each elicitor type for its duration of presentation. The stimuli were masked with back-and-white patterns – and not neutral faces (Kim et al., 2010) – to avoid masked-to-mask stimuli emotional incongruency (masked fearful and happy faces) and congruency (masked neutral faces) perceptual biases (see Tsikandilakis et al., 2022c; pp. 16–19).

(see Kim et al., 2010). Other reviews have argued that neutral faces are recognised less accurately because they lack evolutionary importance (see Cosmides and Tooby, 2000). Some reviews have argued that – as mentioned also in this manuscript – the high spatial characteristics of neutral faces – such as the standard deviation of their luminance – could make them less discernible (see Droit-Volet, Brunot and Niedenthal, 2004). Not surprisingly, a third set of review articles has argued that evolutionary importance and spatial characteristics interact to make neutral faces less discernible than other stimulus types, such as fearful, happy and angry faces (see Hedger et al., 2019). It is still debated why this effect occurs (we are preparing two forthcoming reviews relevant to this subject: Tsikandilakis et al., 2023b; 2023c). As regards the main objective of our analyses, we showed very substantial evidence for the null for all three elicitor types, and we could proceed to the next stage with these durations to test whether individual unconsciousness can lead to self-report and physiological responses

Testing responsivity to unconscious stimuli

For stage two, our calculations suggest that we should present this particular participant with forty 8.33 ms fearful faces, forty 16.67 ms happy faces and forty 25 ms neutral faces. This also means that we will have to present forty 8.33 ms Gaussian blurs, forty 16.67 ms Gaussian blurs and forty 25 ms Gaussian blurs. The resulting design is illustrated in Fig. 3.

Our next subject topic is assessment. We would like to assess the physiology of the participants and also their self-report ratings in response to our individually adjusted stimuli. For physiological assessments, we have a certain set of combined methods that we have used in past studies (see Tsikandilakis et al., 2020a). We have used skin conductance (SCR), heart rate (HR) and facial-emotional expression recognition (FE) for our physiological assessments. Our assessments involve SCR measured from the left hand (index/first and middle/second fingers) of each participant using disposable Ag/AgCl gelled electrodes received by a BIOPAC System, EDA100C in units of micro-Siemens (μS) and recorded in AcqKnowledge. The presence of a phasic skin conductance response is defined as an unambiguous increase (0.01 μS) with respect to each pre-target SCR score occurring one to three seconds post-elicitor offset. The presence of a heart-rate response is defined as an event-related heart-rate peak in beats per minute with respect to each pre-target heart-rate score occurring one to five seconds post-elicitor offset.

For facial expressions analyses, we use Noldus Face-Reader versions 7.1 to 9.1 (see Sabatos-DeVito et al., 2019). We use an HD (4 K) camera mounted at the bottom of the presenting screen and centered on the participant's face. The analysis is run using the maximum video capture frames per second allowed by the face-reader equipment (30 fps). It includes a custom template, and each participant is evaluated in respect to an expressed emotion after controlling for facial-emotional characteristics that are present in their neutral expressions using the participant calibration module. Expression of an emotion is defined as recognition of an emotional expression up to five seconds post-elicitor offset (Tsikandilakis et al., 2020a, 2020b; see also Skiendziel et al., 2019) .[4]

[4] These are the psychophysiological assessments that we have used in previous publications. If the reader would like to apply additional or different psychophysiological assessments, relevant procedures can be found in Cacioppo, Tassinary and Berntson (2007; pp. 19-68 (fMRI & fNIRS); 85-91 (EEG); 120-131 (TMS); 211-218 (Gastrointestinal Responses); 231-235 (Respiratory Responses); see also Tsikandilakis and colleagues (2020c)). In the aforementioned volume the reader will also be able to find additional processes and parameters for measuring SCR (pp. 159-163), HR (pp. 182-197) and facial responses (pp. 267-291).

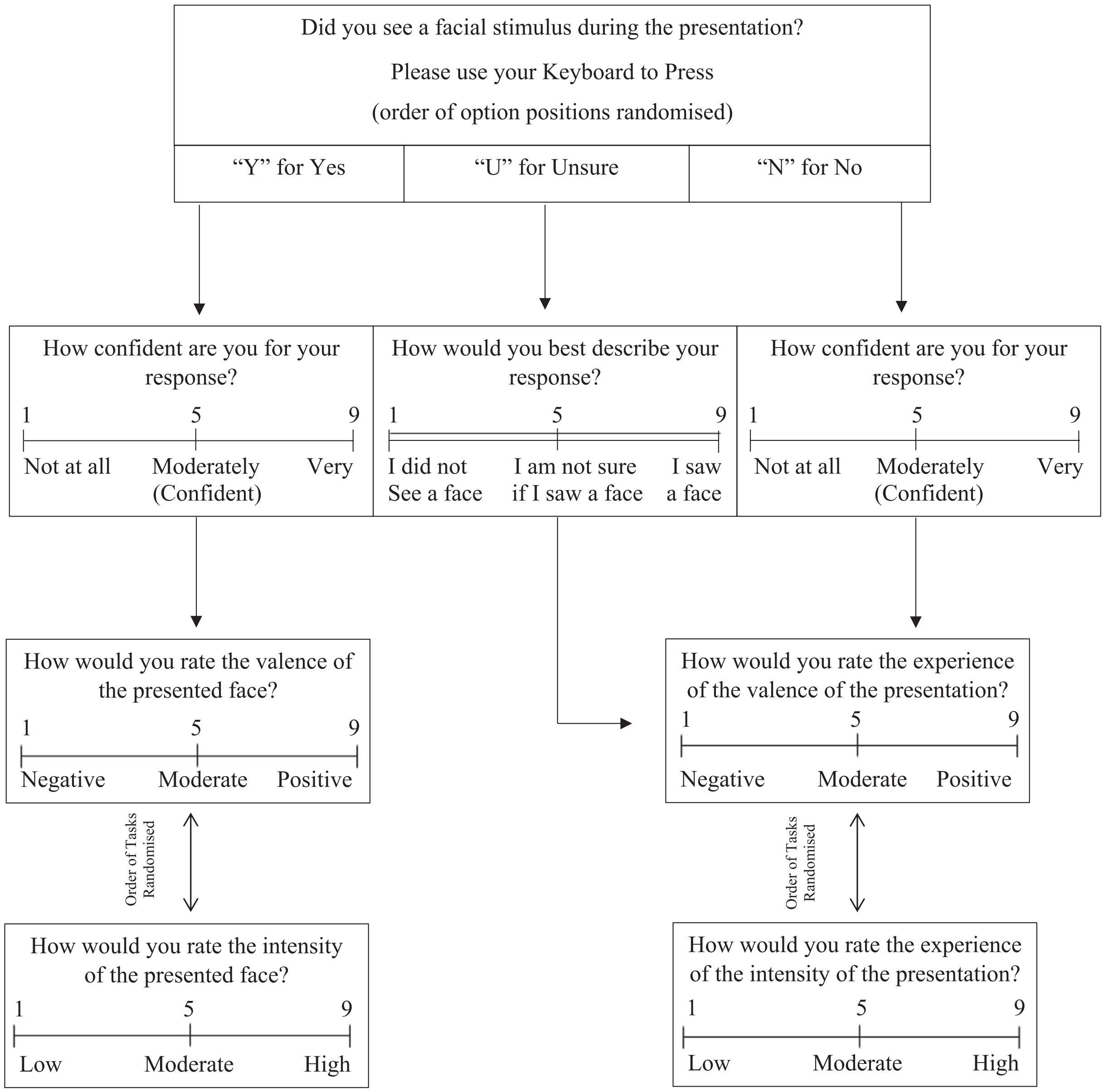

Fig. 4. Engagement Task Matrix.
Matrix of conditional branching responses in each trial following the stimulus presentation and the window of assessment for psychophysiological responses. This setup allows us to infer specific values for each response type such as hits, seeing a face that was presented, misses, not seeing a face that was presented, and uncertainty responses, not being able to conclusively answer whether the participant saw or not a face that was presented. Significant effects for physiological changes and self-report ratings in response to emotional faces for the latter two conditions could stand as evidence that partial awareness and/or unconscious effects were reported using unbiased adjustments for individual unconsciousness. We use the word "very" as opposed to extremely in Likert scales because our pilot data have repeatedly shown that the latter deters the participants from scores ≥ 7, therefore, biasing the design towards lesser ratings. We use the word "moderate" as opposed to "neutral" in Likert scales relating to valence and intensity to maintain homogeneity between different engagement tasks (see Tsikandilakis et al., 2019, 2020a, 2021a)

For these assessments to provide us with meaningful results we need to develop a matrix of post-trial and post-physiological assessment functions. This should assess hits, misses and it will be wise to include uncertainty as an indication of non-categorical conscious perception and unawareness of the presented elicitors (see Morin, 2006). We could also benefit from confidence ratings for these responses, particularly in the interest of linear trend analyses (see particularly Tsikandilakis et al., 2018; pp. 79–83) and ratings for valence and intensity for every post-trial set of engagement tasks (Tsikandilakis & Chapman, 2018; Tsikandilakis et al., 2018, 2019b, 2020a; see Figure 4; for the coding script presented in Fig. 4, see https://osf.io/xy9ru/).

We are ready now to proceed to the method for our analyses and – perhaps, most interestingly – our outcomes.[5] The usual approach in relevant research is to run a repeated-measures ANOVA with independent variable Elicitor Type (fear, happy & neutral) and dependent variable – for the purposes of this illustration – SCR (see Critchley, 2002; Mertens and Engelhard, 2020). In case of significant results for the omnibus ANOVA, and significant differences between Bonferroni-corrected

[5] We will use three decimal points in this section to illustrate more accurately to the reader the relevant results because the latter often involve very discrete and small values and differences.

elicitor type comparisons, such as higher physiological arousal in response to fearful faces compared to happy and neutral faces, our outcomes could be interpreted as evidence for unconscious emotional responsivity. They should not (Tsikandilakis et al., 2019b; 2020a).

To make such inferences, we need to know where the responses are coming from. We need to know what the responses of the participants were to faces they reported seeing when these faces were presented (hits). We need to know what the responses of the participants were to faces they reported they were unsure whether they saw or not when these faces were presented (uncertainty). We need to know what the responses of the participants were to faces they reported that they did not see when they were presented (misses).

To illustrate the importance of the division mentioned above, in our aforementioned 2020 pilot, we encountered a finding that we did not expect and was not in accordance with the vast majority of our research outcomes when we previously applied individual unconsciousness (see Tsikandilakis and Chapman, 2018; pp. 443–449; Tsikandilakis et al., 2018; pp. 87–91; Tsikandilakis et al., 2019b; pp. 21–25; Tsikandilakis et al., 2020a; pp. 501–507; Tsikandilakis et al., 2021; pp. 13–18; Tsikandilakis et al., 2022c; pp. 23–29). A repeated measures ANOVA with independent variable Elicitor Type (fear, happy & neutral) and dependent variable skin conductance responses showed a significant effect of Elicitor Type (F (1.642, 83.747) = 371.249; $p < .001$; $\eta^2_p = 0.879$; Greenhouse-Geisser corrected; SE = 0.01; BF = + ∞). Further Bonferroni-corrected comparisons showed that fearful faces elicited higher SCR ($M = 0.041$; SD = 0.005) compared to happy ($M = 0.028$; SD = 0.004; $p < .001$; $d = 2.87$; SE = 0.001; BF = + ∞) and neutral faces ($M = 0.017$; SD = 0.003; $p < .001$; $d = 5.82$; SE = 0.001; BF = + ∞). Happy faces were also higher for SCR compared to neutral faces ($p < .001$; $d = 3.11$; SE = 0.01; BF = + ∞). We have significant results for unconscious emotional responsivity as reported by higher SCR for fearful faces compared to happy and neutral faces. In fact, we could argue that happy faces also provided evidence for unbiased unconscious emotional responsivity if their comparison to neutral faces can suffice to justify this claim (Winkielman and Berridge, 2004).

It is accurate to say that hits, reporting seeing a masked face when a masked face was presented, can be referred to as conscious perception. On the other hand, miss responses, such as reporting not seeing a masked face when a masked face was presented, can be referred to as a presentation in which participants were not conscious of the masked target. The responses for uncertainty could correspond between levels one, “almost no experience at all, any responses would reflect mere guesses” and two, “a brief glimpse, a feeling that something has been shown that cannot be characterized by any content and that cannot be explained further”, on the PAS scale (Sandberg and Overgaard, 2015; pp. 182–185; see also Overgaard and Sandberg, 2021). It would be fair to call this condition of perception partial between consciousness and unconsciousness, or simply perceptual uncertainty.

In our pilot, a repeated-measures ANOVA for consciously perceived stimuli (hits), with the same independent and dependent variables described above, gives us significant results (F (1.449, 73.877) = 421.708; $p < .001$; $\eta^2_p = 0.892$; Greenhouse-Geisser corrected; SE = 0.01; BF = + ∞). Further Bonferroni-corrected comparisons reveal higher SCR for hits for fearful faces ($M = 0.069$; SD = 0.011) compared to hits for happy ($M = 0.041$; SD = 0.011; $p < .001$; $d = 2.545$; SE = 0.001; BF = + ∞) and hits for neutral faces ($M = 0.017$; SD = 0.004; $p < .001$; $d = 6.282$; SE = 0.001; BF = + ∞). Hits for happy faces were higher for SCR than hits neutral faces ($p < .001$ $d = 2.899$; SE = 0.001; BF = + ∞). The same analysis for unconsciousness (miss responses) presents us with very different outcomes. No significant differences and Bayesian evidence for equivalence of significance for the null were reported between Elicitor Types (F (2, 102) = 1.767; $p = .176$; $\eta^2_p = 0.033$; SE = 0.001; BF = 0.31). We can provide further Bonferroni-corrected comparisons, and these also confirm that SCR were not significantly different and provided evidence for the null for miss responses for fearful faces ($M = 0.017$; SD = 0.005) compared to miss responses for happy ($M = 0.017$; SD = 0.005; $p = 534$; $d = 0.007$; SE = 0.001; BF = 0.18) and miss responses to neutral faces ($M = 0.017$; SD = 0.004; $p = .229$; $d = 0.025$; SE = 0.001; BF = 0.27), and miss responses for happy faces compared to miss responses for neutral faces ($p = .999$; $d = 0.002$; SE = 0.001; BF = 0.03).

For perceptual uncertainty, significant differences were reported between elicitor types (F (2, 102) = 84.224; $p < .001$; $\eta^2_p = 0.623$; SE = 0.01; BF = + ∞). Responses for uncertainty when fearful faces were presented ($M = 0.037$; SD = 0.009) were higher for SCR compared to responses for uncertainty when happy ($M = 0.027$; SD = 0.008; $p < .001$; $d = 1.174$; SE = 0.001; BF = + ∞) and neutral faces ($M = 0.018$; SD = 0.005; $p < .001$; $d = 2.609$; SE = 0.001; BF = + ∞) were presented. Responses for uncertainty when happy faces were presented were higher for SCR than responses for uncertainty when neutral faces ($p < .001$; $d = 1.236$; SE = 0.01; BF = + ∞). Perceptual uncertainty provided evidence that participants experienced higher electrodermal responses to fearful faces compared to any other facial elicitor, and for happy faces compared to neutral faces.

These findings illustrate our method and are important (see Brooks et al., 2012). Conscious perception (hits) showed significant physiological responsivity, unconsciousness (miss responses) showed evidence for the null, and perceptual uncertainty provided us with significant findings for differences in skin conductance responses between fearful, happy and neutral faces. Our method was illustrated, applied and – in this instance – resulted in quite unexpected findings (see Morin, 2006; Ramsøy et al., 2012; Newell and Shanks, 2014; Fisk, and Haase, 2020).

Overall summary

We presented a step-by-step theoretical rationale and empirical illustration-replication guide for implementing individual unconsciousness. We presented the statistical, mathematical and methodological processes for calculating per participant and elicitor type the threshold for chance-level perception. We presented the methodology for assessing self-report ratings and physiological responses to individually adjusted backward masked targets. We presented an empirical illustration for replication including unexpected results for higher arousal for uncertainty between fearful, and happy and neutral faces.

General discussion

Psychological research has been an area of polemic empirical discontent for whether unconscious processing is or is not a real phenomenon. This conflict runs deep and old but so do multiple attempts to improve the methodological canon for assessing whether unconscious processing is real. For example, the implementation of receiver operating characteristics, instead of hit rates, in this area has enabled us to measure perceptual sensitivity without – or at least with less (Stanislaw and Todorov, 1999) – bias and exposure to conservative or liberal participant response strategies and criteria (see Hautus et al., 2021). The implementation of Bayesian analyses has provided us with the means to show direct evidence for the null – that participants were unconscious of a presented target – and has contributed to the assessment of whether participants were unconscious during the presentation of masked elicitors (Dienes, 2014; 2015; 2016). We have added to these a methodological process. This methodological process includes adjustments in the duration of masked presentations. These adjustments are made separately for each participant and for each elicitor type.

This method consists of a first stage during which the durations of presentation that provide evidence for the null hypothesis – that participants were proximate or at-chance level for perceiving a masked target – are selected. It consists of a second stage during which participants are presented with masked targets for these pre-defined durations. Their responses, such as responses for seeing a presented face (hits), responses for not seeing a presented faces (miss responses), and responses for being uncertain of having seen a presented face (responses for uncertainty), are assessed. Their assessments can include self-report ratings and/or physiological measurements, such as SCR, HR and facial-emotional analyses (see Tsikandilakis et al., 2019b). The results are interpreted overall and separately for each response (i.e., hits and miss responses, and responses for uncertainty; see Tsikandilakis et al., 2020a).

This method is presented here in sufficient detail, including open-source material, open-source methodological, mathematical and statistical code, and empirical illustrations, so that it can be directly replicated (Tsikandilakis and Chapman, 2018; Tsikandilakis et al., 2018, 2019b; 2020a, 2021a, 2022c). The method is laborious, it involves strict participant selection and assessment criteria, and it is overall arduous to implement. It is also effective to the extent that the aim of an experiment involves the attainment and assessment of unbiased unconsciousness (see example van der Ploeg et al., 2017; Vadillo et al., 2020, 2021; Heck et al., 2022).

We have presented a difficult but feasible method. This method can provide us with evidence for response effects under conditions of conscious, unconscious – and potentially – partial awareness, such as responses for uncertainty in a post-trial signal detection task. Whether uncertainty signifies unconscious processing will be extensively debated and we do not make any claims concerning its interpretation at this stage given the ongoing debate on the subject. We submitted the blueprints for unbiased individual unconsciousness, and we have made its outcomes available for discourse (Morin, 2006; Ramsøy et al., 2012; Newell and Shanks, 2014; Fisk, and Haase, 2020; Siegel et al., 2022).

Developing the future of individual unconsciousness metrics

In the current manuscript we refer to unconsciousness. Unconsciousness as presented in these pages, refers to unawareness of a presented stimulus under conditions of backward masking. It does necessarily relate to the *zeitgeist* of unconscious behavior in situ – or simply in real everyday life – as presented in relevant literature (see for example, Bargh and Hassin, 2021) which involves whether and to what extent our behavior, intentions, motivations, emotional self-awareness and metacognition, and automatic and involuntary responses stem from unconscious process with or without the involvement of conscious awareness (see Reber and Allen, 2022). These themes remain controversial and a uniform academic perspective as to how to address them is priming studies (Meyen et al., 2022).Therefore, an important – albeit implicit – new challenge we have raised is to extend the current methodology to priming and beyond the limited five-second post-trial physiological and participant engagement-task assessment window following visual suppression (see Bargh and Morsella, 2008).

This subject might seem extraneous but, having delivered the blueprints for the current method, it is only natural to contemplate that in the last 57 years unconscious priming has transformed from a notorious publicity hoax to a methodologically debated empirical reality. Unconscious priming begun empirically with evidence for being a conditional effect, that can take place unconsciously, only when the unconscious prime can satisfy an already existing desire. Recently, it has landed at a point that relevant research has provided evidence that unconscious primes can unconditionally influence – even against our personal ideologies and beliefs – our interpersonal, cultural, racial, political, religious and emotional behavior (see Bar and Biederman, 1998; Karremans et al., 2006; Van Den Bussche et al., 2009; Elgendi et al., 2018; Albarrak et al., 2021; Tsikandilakis et al., 2022a).

These empirical evidence are subject to the limitations that we discoursed in our introduction. They have not been tested using individual thresholds (see Meyen et al., 2022). The future of individual unconsciousness metrics can be to explore these most interesting and most provocative claims. It can be to explore what is the outcome of unbiased individual unconscious presentations for our subsequent perceptions, and cognitive and emotional responses. The next step of this method, now that its rationale and replicating tools have been made available, can be its application within a context of ecological validity, and the exploration of possibly the most enduring myth – or possibly the most unsettling reality – of research relating to the unconscious: Can we be influenced to alter our emotions, cognition and behaviours when presented with truly unconscious cues (Tsikandilakis et al., 2023d; in preparation)?

Conclusions

In this manuscript, we presented the methodological, mathematical, statistical and outcome assessment framework for unbiased individual unconsciousness. We offered open-source material, code and the required resources to replicate our method. We showed in several empirical illustrations that this methodology is laborious in its implementation but that it is efficient in achieving unbiased per participant and elicitor-type chance-level perceptual sensitivity. We provided empirical illustrations. We showed challenging results for higher emotional responsivity for fearful compared to happy and neutral faces only for responses for consciously perceiving (hits) and responses for partial awareness (uncertainty) of masked elicitors. We emphasized that perceptual uncertainty is a condition that merits discourse. We emphasized that the developing future research applications of individual unconsciousness can be its implementation for undertaking research in priming studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data for this manuscript can be found at https://osf.io/78qbh/.

Acknowledgements

The interested reader or scholar or researcher from any academic discipline or area, and equally so the non-academic reader are very welcome and encouraged to contact the corresponding author - or any other member of our team, as they prefer – for any additional information they would like to discuss concerning the current method, relating to its replication, specific code and mathematical scripts, or simply concerning how and their opinion on whether this method works. We warmly welcome all correspondence. All included studies were approved by the Ethics Committee of the School of Psychology of the University of Nottingham. All data, and experimental, statistical and mathematical code have been made available at https://osf.io/78qbh/.

References

Adolph, D., Alpers, G.W., 2010. Valence and arousal: a comparison of two sets of emotional facial expressions. Am. J. Psychol. 123 (2), 209–219.
Albarrak, L., Metatla, O., Roudaut, A., 2021. Exploring the influence of subliminal stimulus type and peripheral angle on the priming effect. Int. J. Hum. Comput. Stud. 151, 102631.
Albrecht, T., Klapötke, S., Mattler, U., 2010. Individual differences in metacontrast masking are enhanced by perceptual learning. Conscious. Cogn. 19 (2), 656–666.
Albrecht, T., Mattler, U., 2012. Individual differences in subjective experience and objective performance in metacontrast masking. J. Vis. 12 (5), 5 -5.
Arnoldo, T., Víctor, C.V., 2015. Effect size, confidence intervals and statistical power in psychological research. Psychol. Russia 8 (3), 27–46.
Axelrod, V., Bar, M., Rees, G., 2015. Exploring the unconscious using faces. Trends Cogn. Sci. 19 (1), 35–45.
Baker, D.H., Vilidaite, G., Lygo, F.A., Smith, A.K., Flack, T.R., Gouws, A.D., Andrews, T.J., 2020. Power contours: optimising sample size and precision in experimental psychology and human neuroscience. Psychol. Methods.
Balconi, M., Lucchiari, C., 2007. Consciousness and emotional facial expression recognition: subliminal/supraliminal stimulation effect on N200 and P300 ERPs. J. Psychophysiol. 21 (2), 100–108.
Banks, W.P., 1970. Signal detection theory and human memory. Psychol. Bull. 74 (2), 81–93.
Bar, M., Biederman, I., 1998. Subliminal visual priming. Psychol. Sci. 9 (6), 464–468.
Bargh, J.A., Morsella, E., 2008. The unconscious mind. Perspect. Psychol. Sci. 3 (1), 73–79.

Bargh, J.A., Hassin, R.R., 2021. Human unconscious processes in situ: the kind of awareness that really matters. Cognit. Unconscious.
Barrett, L.F., 2006. Solving the emotion paradox: categorization and the experience of emotion. Personal. Soc. Psychol. Rev. 10 (1), 20–46.
Barrett, L.F., Tugade, M.M., Engle, R.W., 2004. Individual differences in working memory capacity and dual-process theories of the mind. Psychol. Bull. 130 (4), 553.
Bayes, T., Hume, D., 1763. Bayes's theorem. Proc. Br. Acad. 113, 91–109.
Berger, J.O., Sellke, T., 1987. Testing a point null hypothesis: the irreconcilability of p values and evidence. J. Am. Stat. Assoc. 82 (397), 112–122.
Berkovits, I., Hancock, G.R., Nevitt, J., 2000. Bootstrap resampling approaches for repeated measure designs: relative robustness to sphericity and normality violations. Educ. Psychol. Meas. 60 (6), 877–892.
Bernardo, J.M., 1980. A Bayesian analysis of classical hypothesis testing. Trabajos de Estadística y de Investigación Operativa 31 (1), 605–647.
Berryman, C., McAuley, J.H., Moseley, L.G., 2012. Sphere 12 screening questionnaire. J. Physiother. 58 (4), 273.
Bishop, S.J., 2008. Neural mechanisms underlying selective attention to threat. Ann. N. Y. Acad. Sci. 1129 (1), 141–152.
Brand, A., Bradley, M.T., Best, L.A., Stoica, G., 2010. Multiple trials may yield exaggerated effect size estimates. J. Gen. Psychol. 138 (1), 1–11.
Breitmeyer, B.G., Ogmen, H., 2000. Recent models and findings in visual backward masking: a comparison, review, and update. Percept. Psychophys. 62 (8), 1572–1595.
Besken, M., Mulligan, N.W., 2013. Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. Mem. Cognit. 41, 897–903.
Brooks, S.J., Savov, V., Allzén, E., Benedict, C., Fredriksson, R., Schiöth, H.B., 2012. Exposure to subliminal arousing stimuli induces robust activation in the amygdala, hippocampus, anterior cingulate, insular cortex and primary visual cortex: a systematic metaanalysis of fMRI studies. Neuroimage 59 (3), 2962–2973.
Burnham, J.C., 1967. Psychoanalysis and american medicine, 1894-1918: medicine, science, and culture. Psychol. Issues.
Cacioppo, J.T., Tassinary, L.G., Berntson, G. (Eds.), 2007. Handbook of Psychophysiology. Cambridge university press.
Carlson, J.M., Reinke, K.S., LaMontagne, P.J., Habib, R., 2011. Backward masked fearful faces enhance contralateral occipital cortical activity for visual targets within the spotlight of attention. Soc. Cogn. Affect. Neurosci. 6 (5), 639–645.
Chambers, C., 2019. The seven deadly sins of psychology. The Seven Deadly Sins of Psychology. Princeton University Press.
Churchman, C.W., 1946. Philosophical aspect of statistical theory. Philos. Rev. 55 (1), 81–87.
Cohen, J., MacWhinney, B., Flatt, M., Provost, J., 1993. PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behav. Res. Methods Instrum. Comput. 25 (2), 257–271.
Cosmides, L., Tooby, J., 2000. Evolutionary psychology and the emotions. Handb. Emot. 2 (2), 91–115.
Cowen, A.S., Keltner, D., 2020. What the face displays: mapping 28 emotions conveyed by naturalistic expression. Am. Psychol. 75 (3), 349–361.
Critchley, H.D., 2002. Electrodermal responses: what happens in the brain. Neuroscientist 8 (2), 132–142.
De Gardelle, V., Kouider, S., 2010. How spatial frequencies and visual awareness interact during face processing. Psychol. Sci. 21 (1), 58–66.
Dienes, Z., 2014. Using Bayes to get the most out of non-significant results. Front. Psychol. 5, 781–795.
Dienes, Z., 2015. How Bayesian statistics are needed to determine whether mental states are unconscious. In: Overgaard, Morten (Ed.), Behavioural Methods in Consciousness Research. Oxford University Press, Oxford, pp. 199–220.
Dienes, Z., 2016. How Bayes factors change scientific practice. J. Math. Psychol. 72, 78–89.
Dienes, Z., 2019. How do I know what my theory predicts? Adv. Methods Pract. Psychol. Sci. 2, 364–377.
Dienes, Z., 2021. How to use and report Bayesian hypothesis tests. Psychol. Consciousness 8 (1), 9–26.
Dienes, Z., Coulton, S., Heather, N., 2018. Using Bayes factors to evaluate evidence for no effect: examples from the SIPS project. Addiction 113 (2), 240–246.
Dixon, N.F., 1971. Subliminal Perception: The Nature of a Controversy. McGraw-Hill, London.
Dixon, N., 1981. Preconscious Processing. Wiley, Toronto.
Donges, U.S., Kersting, A., Suslow, T., 2014. Alexithymia and perception of emotional information: a review of experimental psychological findings. Univ. Psychol. 13 (2), 745–756.
Droit-Volet, S., Brunot, S., Niedenthal, P., 2004. BRIEF REPORT Perception of the duration of emotional events. Cognit. Emot. 18 (6), 849–858.
Ebbinghaus, H., 1908. Psychology: An elementary Text-Book. DC Heath Press.
Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference for psychological research. Psychol. Rev. 70 (3), 193–198.
Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., Cichocki, A., 2018. Subliminal priming—State of the art and future perspectives. Behav. Sci. 8 (6), 54–66.
Erdelyi, M.H., 2004. Subliminal perception and its cognates: theory, indeterminacy, and time. Conscious. Cogn. 13 (1), 73–91.
Faul, F., Erdfelder, E., Buchner, A., Lang, A.G., 2009. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41 (4), 1149–1160.
Fechner, G.T., 1948. Elements of psychophysics, 1860. In: Dennis, W. (Ed.), Readings in the History of Psychology. Appleton-Century-Crofts, pp. 206–213.
Field, G.C., Aveling, F., Laird, J., 1922. Is the conception of the unconscious of value in psychology. Mind 31 (124), 413–442.
Fisk, G.D., Haase, S.J., 2020. Binary vs. continuous experimental designs for the study of unconscious perceptual processing. Conscious. Cogn. 81, 102933.
Frosh, S., 1989. Psychoanalysis and Psychology: Minding the Gap. Macmillan International Higher Education.
Garrido, L.E., Abad, F.J., Ponsoda, V., 2016. Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. Psychol. Methods 21 (1), 93.
Garrido, A., Duschek, S., Árbol, J.R., Usera, I.G., Vila, J., Mata, J.L., 2020. Sympathetic contributions to habituation and recovery of the cardiac defense response. Biol. Psychol. 151, 107846.
Goldiamond, I., 1958. Indicators of perception: I. Subliminal perception, subception, unconscious perception: an analysis in terms of psychophysical indicator methodology. Psychol. Bull. 55 (6), 373.
Goodkin, O., Phillips, M.A., 1980. The subconscious taken captive: a social, ethical, and legal analysis of subliminal communication technology. S. Cal. L. Rev. 54, 1077.
Gray, K.L., Adams, W.J., Hedger, N., Newton, K.E., Garner, M., 2013. Faces and awareness: low-level, not emotional factors determine perceptual dominance. Emotion 13 (3), 537.
... & Gross, A.L., Inouye, S.K., Rebok, G.W., Brandt, J., Crane, P.K., Parisi, J.M., Jones, R.N., 2012. Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. J. Clin. Exp. Neuropsychol. 34 (7), 758–772.
Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F., 1998. Making faces. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 55–66.
Gullone, E., Taffe, J., 2012. The emotion regulation questionnaire for children and adolescents (ERQ–CA): a psychometric evaluation. Psychol. Assess. 24 (2), 409.
Gur, R.C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., … Gur, R.E., 2002. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. J. Neurosci. Methods 115 (2), 137–143.
Heck, D.W., Boehm, U., Böing-Messing, F., Bürkner, P.C., Derks, K., Dienes, Z., … Hoijtink, H., 2022. A review of applications of the bayes factor in psychological research. Psychol. Methods.
Hedger, N., Adams, W.J., Garner, M., 2015. Fearful faces have a sensory advantage in the competition for awareness. J. Exp. Psychol. 41 (6), 1748–1759.
Hedger, N., Gray, K.L., Garner, M., Adams, W.J., 2016. Are visual threats prioritized without awareness? A critical review and meta-analysis involving 3 behavioral paradigms and 2696 observers. Psychol. Bull. 142 (9), 934–957.
Hedger, N., Garner, M., Adams, W.J., 2019. Do emotional faces capture attention, and does this depend on awareness? Evidence from the visual probe paradigm. J. Exp. Psychol. 45 (6), 790–802.
Hess, U., Kafetsios, K., Mauersberger, H., Blaison, C., Kessler, C.L., 2016. Signal and noise in the perception of facial emotion expressions: from labs to life. Personal. Soc. Psychol. Bull. 42 (8), 1092–1110.
Hickie, I.B., Davenport, T.A., Hadzi-Paviovic, D., Koschera, A., Naismith, S.L., Scott, E.M., Wilhelm, K.A., 2001. Development of a simple screening tool for common mental disorders in general practice. Medical J. Austral. 175, S10–S17.
Howard, G.S., Maxwell, S.E., Fleming, K.J., 2000. The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. Psychol. Methods 5 (3), 315–331.
Hautus, M.J., Macmillan, N.A., Creelman, C.D., 2021. Detection Theory: A User's Guide. Routledge.
Haynes, S.N., Lench, H.C., 2003. Incremental validity of new clinical assessment measures. Psychol. Assess. 15 (4), 456–461.
Izmailov, C.A., Chernorizov, A.M., Polyansky, V.B., 2022. Human–Neuron–Model": a Spherical Model of Signal Discrimination in the Visual System. In: Russian Cognitive Neuroscience. Brill, pp. 154–177.
Japee, S., Crocker, L., Carver, F., Pessoa, L., Ungerleider, L.G., 2009. Individual differences in valence modulation of face-selective M170 response. Emotion 9 (1), 59–67.
Kahn, E., 1943. Unconsciousness. Am. J. Psychiatry 99 (5), 771 -771.
Karremans, J.C., Stroebe, W., Claus, J., 2006. Beyond Vicary's fantasies: the impact of subliminal priming and brand choice. J. Exp. Soc. Psychol. 42 (6), 792–798.
Kaspar, K., König, P., 2012. Emotions and personality traits as high-level factors in visual attention: a review. Front. Hum. Neurosci. 6, 321–341.
Kim, M.J., Loucks, R.A., Neta, M., Davis, F.C., Oler, J.A., Mazzulla, E.C., Whalen, P.J., 2010. Behind the mask: the influence of mask-type on amygdala response to fearful faces. Soc. Cogn. Affect. Neurosci. 5 (4), 363–368.
Kratochwill, T.R., Levin, J.R, 2014. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. In: Kratochwill, T.R., Levin, & J.R. (Eds.), Single-case Intervention research: Methodological and Statistical Advances. American Psychological Association, pp. 53–89. doi:10.1037/14376-003.
Kring, A.M., Sloan, D.M., 2007. The Facial Expression Coding System (FACES): development, validation, and utility. Psychol. Assess. 19 (2), 210.
Kruschke, J.K., 2011. Bayesian assessment of null values via parameter estimation and model comparison. Perspect. Psychol. Sci. 6 (3), 299–312.
Kruschke, J.K., Liddell, T.M., 2018. The Bayesian New Statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon. Bull. Rev. 25 (1), 178–206.
Lane, D., 2016. The assumption of sphericity in repeated-measures designs: what it means and what to do when it is violated. Quant. Methods Psychol. 12, 114–122.
Lann, E., 1959. Testing Statistical Hypotheses. Wiley, New York.
Leong, M.Q., Yu, Z., Tsikandilakis, M., Tong, E.M.W., 2022. "See no Evil. Feel no Evil?": exploring emotional responses to masked moral violations in religious and non-religious singaporean participants. Evolut. Behav. Sci..
Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49 (4), 764–766.

Li, Y., Wright, D.L., 2000. An assessment of the attention demands during random-and blocked-practice schedules. Quart. J. Exp. Psychol. Sect. A 53 (2), 591–606.

Lindley, D.V., 1957. A statistical paradox. Biometrika 44 (1/2), 187–192.

Lord, R.G., 1985. Accuracy in behavioral measurement: an alternative definition based on raters' cognitive schema and signal detection theory. J. Appl. Psychol. 70 (1), 66–74.

Madipakkam, A.R., Rothkirch, M., 2019. The unconscious processing of social information. In: Hesselmann, G. (Ed.), Transitions Between Consciousness and Unconsciousness. Routledge/Taylor & Francis Group, pp. 92–117.

Macmillan, N.A., 1993. Signal detection theory as data analysis method and psychological decision model. In: Keren, G., Lewis, C. (Eds.), A Handbook For Data Analysis in the Behavioral Sciences: Methodological Issues. Lawrence Erlbaum Associates, Inc, pp. 21–57.

Maldonato, N.M., Muzii, B., Bottone, M., Sperandeo, R., Di Corrado, D., Continisio, G.I., … Esposito, A., 2020. Unitas Multiplex. Biological architectures of consciousness. Infocommun. J. 12 (1), 10–16.

Maniscalco, B., Lau, H., 2014. Signal detection theory analysis of type 1 and type 2 data: meta-d′, response-specific meta-d′, and the unequal variance SDT model. In: The Cognitive Neuroscience of Metacognition, pp. 25–66.

Maxwell, S.E., Lau, M.Y., Howard, G.S., 2015. Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? Am. Psychol. 70 (6), 487–499.

Mauss, I.B., Levenson, R.W., McCarter, L., Wilhelm, F.H., Gross, J.J., 2005. The tie that binds? Coherence among emotion experience, behavior, and physiology. Emotion 5 (2), 175 -101.

Meneguzzo, P., Tsakiris, M., Schioth, H.B., Stein, D.J., Brooks, S.J., 2014. Subliminal versus supraliminal stimuli activate neural responses in anterior cingulate cortex, fusiform gyrus and insula: a meta-analysis of fMRI studies. BMC Psychol 2 (1), 52–71.

Meyen, S., Zerweck, I.A., Amado, C., von Luxburg, U., Franz, V.H., 2022. Advancing research on unconscious priming: when can scientists claim an indirect task advantage? J. Exp. Psychol. 151 (1), 65.

Mertens, G., Engelhard, I.M., 2020. A systematic review and meta-analysis of the evidence for unaware fear conditioning. Neurosci. Biobehav. Rev. 108, 254–268.

Merikle, P.M., Cheesman, J., 1987. Current Status of Research on Subliminal Perception. ACR North American Advances.

Miller, J.G, 1942. Unconsciousness. John Wiley & Sons Inc.

Moore, T.E., 2008. Subliminal perception: facts and fallacies. In: Lilienfeld, S.O., Ruscio, J., Lynn, S.J. (Eds.), Navigating the Mindfield: A User's Guide to Distinguishing Science from Pseudoscience in Mental Health. Prometheus Books/Rowman & Littlefield, pp. 589–601.

Morin, A., 2006. Levels of consciousness and self-awareness: a comparison and integration of various neurocognitive views. Conscious. Cogn. 15 (2), 358–371.

Murphy, K.R., Myors, B., Wolach, A., 2014. Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests. Routledge.

Newell, B.R., Shanks, D.R., 2014. Unconscious influences on decision making: a critical review. Behavioral and brain sciences 37 (1), 1–19.

Öhman, A., 2009. Of snakes and faces: an evolutionary perspective on the psychology of fear. Scand. J. Psychol. 50 (6), 543–552.

Ochsner, K., Gross, J.J., 2005. The cognitive control of emotion. Trends Cogn. Sci. (Regul. Ed.) 9, 242–249.

Overgaard, M., Sandberg, K., 2021. The Perceptual Awareness Scale—Recent controversies and debates. Neurosci. Consciousness 2021 (1), niab044.

Pauker, K., Rule, N.O., Ambady, N., 2010. Ambiguity and social perception. In: The Social Psychology of Visual Perception, pp. 13–29.

Pegna, A.J., Darque, A., Berrut, C., Khateb, A., 2011. Early ERP modulation for task-irrelevant subliminal faces. Front. Psychol. 2, 88–97.

Pessoa, L., 2005. To what extent are emotional visual stimuli processed without attention and awareness? Curr. Opin. Neurobiol. 15 (2), 188–196.

Pessoa, L., McKenna, M., Gutierrez, E., Ungerleider, L.G., 2002. Neural processing of emotional faces requires attention. Proc. Natl. Acad. Sci. 99 (17), 11458–11463.

Pessoa, L., Ungerleider, L.G., 2004. Neuroimaging studies of attention and the processing of emotion-laden stimuli. Prog. Brain Res. 144, 171–182.

Pessoa, L., Adolphs, R., 2010. Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. Nat. Rev. Neurosci. 11 (11), 773–782.

Peterson, W., Birdsall, T., Fox, W., 1954. The theory of signal detectability. Trans. IRE Profession. Group Inf. Theory 4 (4), 171–212.

van der Ploeg, M.M., Brosschot, J.F., Versluis, A., Verkuil, B., 2017. Peripheral physiological responses to subliminally presented negative affective stimuli: a systematic review. Biol. Psychol. 129, 131–153.

Pollack, I., Norman, D.A., 1964. A non-parametric analysis of recognition experiments. Psychon. Sci. 1 (1), 125–126.

Ramsøy, T.Z., Friis-Olivarius, M., Jacobsen, C., Jensen, S.B., Skov, M., 2012. Effects of perceptual uncertainty on arousal and preference across different visual domains. J. Neurosci. Psychol. Econ. 5 (4), 212–237.

Rassovsky, Y., Horan, W.P., Lee, J., Sergi, M.J., Green, M.F., 2011. Pathways between early visual processing and functional outcome in schizophrenia. Psychol. Med. 41 (3), 487–497.

Reber, A.S., Allen, R. (Eds.), 2022. The Cognitive Unconscious: The First Half Century. Oxford University Press.

Rosellini, A.J., Brown, T.A., 2021. Developing and validating clinical questionnaires. Annu. Rev. Clin. Psychol. 17, 55–81.

Rosnow, R.L., Rosenthal, R., 2003. Effect sizes for experimenting psychologists. Canad. J. Exp. Psychol. 57 (3), 221–234.

Rouder, J.N., Morey, R.D., Speckman, P.L., Pratte, M.S., 2007. Detecting chance: a solution to the null sensitivity problem in subliminal priming. Psychon. Bull. Rev. 14 (4), 597–605.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16 (2), 225–237.

Rouder, J.N., Haaf, J.M., Vandekerckhove, J., 2018. Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. Psychon. Bull. Rev. 25 (1), 102–113.

Rozeboom, W.W., 1960. The fallacy of the null-hypothesis significance test. Psychol. Bull. 57 (5), 416–428.

Sabatos-DeVito, M., Murias, M., Dawson, G., Howell, T., Yuan, A., Marsan, S., … McPartland, J.C., 2019. Methodological considerations in the use of Noldus EthoVision XT video tracking of children with autism in multi-site studies. Biol. Psychol. 146, 107712.

Sala, M.N., Molina, P., Abler, B., Kessler, H., Vanbrabant, L., van de Schoot, R, 2012. Measurement invariance of the Emotion Regulation Questionnaire (ERQ). A cross-national validity study. Eur. J. Dev. Psychol. 9 (6), 751–757.

Sandberg, K., Overgaard, M., 2015. Using the perceptual awareness scale (PAS). In: Behavioural Methods in Consciousness Research, pp. 181–195.

Schyns, P.G., Oliva, A., 1997. Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. Perception 26 (8), 1027–1038.

Searle, J.R., 1989. Consciousness, unconsciousness, and intentionality. Philos. Top. 17 (1), 193–209.

Shanks, D.R., Malejka, S., Vadillo, M.A., 2021. The challenge of inferring unconscious mental processes. Exp. Psychol. 68 (3), 113 -.

Shapley, R., 1990. Visual sensitivity and parallel retinocortical channels. Annu. Rev. Psychol. 41 (1), 635–658.

Sparfeldt, J.R., Schilling, S.R., Rost, D.H., Thiel, A., 2006. Blocked versus randomized format of questionnaires: a confirmatory multigroup analysis. Educ Psychol Meas 66 (6), 961–974.

Skiendziel, T., Rösch, A.G., Schultheiss, O.C., 2019. Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. PLoS One 14 (10), e0223905.

Smith, W.D., 1995a. Clarification of sensitivity measure A'. J. Math. Psychol. 39 (1), 82–89.

Siegel, P., Cohen, B., Warren, R., 2022. Nothing to fear but fear itself: a mechanistic test of unconscious exposure. Biol. Psychiatry 91 (3), 294–302.

Soto, D., Sheikh, U.A., Rosenthal, C.R., 2019. A novel framework for unconscious processing. Trends Cogn. Sci. 23 (5), 372–376.

Spillmann, L., Werner, J.S. (Eds.), 2012. Visual perception: the Neurophysiological Foundations. Elsevier.

Stafford, T., 2014. The perspectival shift: how experiments on unconscious processing don't justify the claims made for them. Front. Psychol. 5, 1067.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. 31 (1), 137–149.

Stein, T., Utz, V., Van Opstal, F., 2020. Unconscious semantic priming from pictures under backward masking and continuous flash suppression. Conscious. Cogn. 78, 102864.

Swets, J.A., 2014. Signal Detection Theory and ROC Analysis in Psychology and diagnostics: Collected papers. Psychology Press.

Taylor, G.J., Bagby, R.M., 2000. An overview of the alexithymia construct. In: Bar-On, R., Parker, J.D.A. (Eds.), The Handbook of Emotional Intelligence: Theory, Development, Assessment, and Application at Home, School, and in the Workplace. Jossey-Bass, pp. 40–67.

Thompson, J., 2007. Online Alexithymia Questionnaire—G2. Retrieve: oaq. blogspot. com.

Tsikandilakis, M., Chapman, P., 2018. Skin conductance responses to masked emotional faces are modulated by hit rate but not signal detection theory adjustments for subjective differences in the detection threshold. Perception 47 (4), 432–450.

Tsikandilakis, M., Chapman, P., Peirce, J., 2018. Target meta-awareness is a necessary condition for physiological responses to masked emotional faces: evidence from combined skin conductance and heart rate assessment. Conscious. Cogn. 58, 75–89.

Tsikandilakis, M., Bali, P., Chapman, P., 2019. Beauty is in the eye of the beholder: the appraisal of facial attractiveness and its relation to conscious awareness. Perception 48 (1), 72–92.

Tsikandilakis, M., Kausel, L., Boncompte, G., Yu, Z., Oxner, M., Lanfranco, R., … Chapman, P., 2019a. There is no face like home": ratings for cultural familiarity to own and other facial dialects of emotion with and without conscious awareness in a British sample. Perception 48 (10), 918–947.

Tsikandilakis, M., Bali, P., Derrfuss, J., Chapman, P., 2019b. The unconscious mind: from classical theoretical controversy to controversial contemporary research and a practical illustration of the "error of our ways. Conscious. Cogn. 74-92, 102771.

Tsikandilakis, M., Bali, P., Derrfuss, J., Chapman, P., 2020a. I can see you; I can feel it; and vice-versa": consciousness and its relation to emotional physiology. Cogn. Emot. 34 (3), 498–510.

Tsikandilakis, M., Bali, P., Derrfuss, J., Chapman, P., 2020b. Anger and hostility: are they different? An analytical exploration of facial-expressive differences, and physiological and facial-emotional responses. Cogn. Emot. 34 (3), 581–595.

Tsikandilakis, M., Bali, P., Haralabopoulos, G., Derrfuss, J., Chapman, P., 2020c. "The Harder One Tries…": findings and insights from the application of covert response pressure assessment technology in three studies of visual perception. Iperception 11 (2), 2041669520913319.

Tsikandilakis, M., Bali, P., Yu, Z., Madan, C., Derrfuss, J., Chapman, P., Groeger, J., 2021a. Individual conscious and unconscious perception of emotion: theory, methodology and applications. Conscious. Cogn. 94, 103172.

Tsikandilakis, M., Yu, Z., Kausel, L., Boncompte, G., Lanfranco, R.C., Oxner, M., … Chapman, P., 2021b. "There is No (Where a) Face Like Home": recognition and appraisal responses to masked facial dialects of emotion in four different national cultures. Perception 50 (12), 1027–1055.

Tsikandilakis, M., Leong, M.Q., Yu, Z., Paterakis, G., Bali, P., Derrfuss, J., … Mitchell, P., 2022a. "Speak of the Devil… and he Shall Appear": religiosity, unconsciousness, and the effects of explicit priming in the misperception of immorality. Psychol. Res. 1–29.

Tsikandilakis, M., Bali, P., Lanfranco, R., Kausel, L., Yu, Z., Boncompte, G., Alshammari, A., Li, R., Milbank, A., Burdett, M., Mével, A., Madan, C., Derrfuss, J., 2022b. Learning emotional dialects: a british population study of cross-cultural communication. Perception doi:10.17605/OSF.IO/SYVF9.

Tsikandilakis, M., Bali, P., Karlis, A., Mével, P., A, Madan, C., Derrfuss, J., Milbank, A, 2022c. Conscious researchers in unconscious research": a motivational revisit of issues, resolutions, and applied awareness in backward masking using faces. Psychol. Rep..

Tsikandilakis, M., Bali, P., Yu, Z., Karlis, A.K., Tong, E.M.W., Milbank, A., Mével, A., Derrfuss, J., Madan, C., 2023a. "The Many Faces of Sorrow": an empirical exploration of the psychological plurality of sadness. Curr. Psychol..

Tsikandilakis, M., Bali, P., Karlis, A., Mével, P., Madan, A., Derrfuss, C., J, Milbank, A, 2023b. "Not My (Stimulus) Type": An Empirical Exploration of Low Perceptual Performance Stimuli Under Conditions of Backward Masking Manuscript in preparation.

Tsikandilakis, M., Bali, P., Karlis, A., Mével, P., A., Madan, C., Derrfuss, J. & Milbank, A. (2023c). *"Sentio Ergo Est": the Case and Correlates of False Positive Responses for Fearful Faces under Static and Individually Adjusted Conditions of Backward Masking.* Manuscript in preparation.

Tsikandilakis, M., Bali, P., Karlis, A., Mével, P., A., Madan, C., Derrfuss, J. & Milbank, A. (2023d). *A Primer on Unconscious Priming: myth or Reality?* Manuscript in preparation.

Vadillo, M.A., Linssen, D., Orgaz, C., Parsons, S., Shanks, D.R., 2020. Unconscious or underpowered? Probabilistic cuing of visual attention. J. Exp. Psychol. 149 (1), 160.

Vadillo, M.A., Malejka, S., Lee, D.Y., Dienes, Z., Shanks, D.R., 2021. Raising awareness about measurement error in research on unconscious mental processes. Psychon. Bull. Rev. 1–23.

delValle, M.V., Andres, M.L., Urquijo, S., Zamora, E.V., Mehta, A., Gross, J.J., 2021. Argentinean adaptation and psychometric properties of the emotion regulation questionnaire (ERQ). Psychol. Rep., 00332941211021343.

Van Den Bussche, E., Van den Noortgate, W., Reynvoet, B., 2009. Mechanisms of masked priming: a meta-analysis. Psychol. Bull. 135 (3), 452–471.

Van Humbeeck, G., Van Audenhove, C., De Hert, M., Pieters, G., Storms, G., 2002. Expressed emotion: a review of assessment instruments. Clin. Psychol. Rev. 22 (3), 321–341.

Verde, M.F., Macmillan, N.A., Rotello, C.M., 2006. Measures of sensitivity based on a single hit rate and false alarm rate: the accuracy, precision, and robustness of′, A z, andA. Percept. Psychophys. 68 (4), 643–654.

Wagenmakers, E.J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R.D., 2018. Bayesian inference for psychology. Part II: example applications with JASP. Psychon. Bull. Rev. 25 (1), 58–76.

Wiggins, B.J., Christopherson, C.D., 2019. The replication crisis in psychology: an overview for theoretical and philosophical psychology. J. Theor. Philos. Psychol. 39 (4), 202–219.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F., Tanaka, J.W., 2010. Controlling low-level image properties: the SHINE toolbox. Behav. Res. Methods 42 (3), 671–684.

Winkielman, P., Berridge, K.C., 2004. Unconscious emotion. Curr. Direct. Psychol. Sci. 13 (3), 120–123.

Wyart, V., Tallon-Baudry, C., 2009. How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. J. Neurosci. 29 (27), 8715–8725.

Yaremko, R.M., Harari, H., Harrison, R.C., Lynn, E., 2013. Handbook of Research and Quantitative Methods in psychology: For students and Professionals. Psychology Press, London.

Yonelinas, A.P., Parks, C.M., 2007. Receiver operating characteristics (ROCs) in recognition memory: a review. Psychol. Bull. 133 (5), 800–817.

Zednik, C., Jakel, F., 2014. How does Bayesian reverse-engineering work? In: Proceedings of the Annual Meeting of the Cognitive Science Society, 36.

Zednik, C., Jäkel, F., 2016. Bayesian reverse-engineering considered as a research strategy for cognitive science. Synthese 193 (12), 3951–3985.

Zhang, J., Mueller, S.T., 2005. A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrika 70, 203–212.