Social Cognition, Vol. 42, No. 1, 2024, pp. 27–60
 https://doi.org/10.1521/soco.2024.42.1.27

IGNORANCE IS NOT BLISS: ON ISSUES MEASURING THE AWARENESS OF SUBOPTIMAL STIMULI

David S. March
Department of Psychology, Florida State University

It is standard practice to assess participants' perception of suboptimal stimuli by using an awareness measure. Yet the assessment of stimulus awareness is a difficult issue in masked priming studies; there is no standard for what constitutes participants' conscious "awareness" nor what measure is best to assess awareness. Nonetheless, researchers make claims of participant (un)awareness based on idiosyncratic operationalizations of "awareness" and unstandardized practices for testing awareness. This unstandardized practice can lead to spurious conclusions based on faulty assumptions. The current work adds to an ongoing discussion on the methodology of the field by drawing attention to how operational definitions and tasks impact the results obtained from experiments. The concept of awareness is briefly discussed, work testing awareness across three attempts is presented, each using different oft-employed awareness measures that render different empirical conclusions, and finally the article discusses choosing an awareness measure that reflects one's research goal.

Keywords: detection, forced-choice identification, subliminal processing, awareness, masking

> There must be a happy medium somewhere between being totally informed and blissfully unaware.
>
> —Doug Larson, columnist and editor

INTRODUCTION

That information processed outside awareness can affect downstream perceptions, judgments, and behaviors is a central topic of study in many fields of psychology. Seminal works in, for example, social (Fazio, 2001; Greenwald et al., 1996), emotional (Zajonc, 1980), and cognitive psychology (Kihlstrom, 1987) have laid the foundation for a vast literature on the consequences of nonconscious processes.

Supplemental materials are available online.

Address correspondence to David S. March, Florida State University, 1107 W. Call St., Tallahassee, FL 32304. E-mail: march@psy.fsu.edu

Yet recent research casts doubt on this position, arguing that the special processing of certain stimuli should not be regarded as settled science (Baier et al., 2022; Hedger et al., 2016; Lohse & Overgaard, 2019; Tsikandilakis & Chapman, 2018). Regardless of which side of the debate one takes, central to this research is *minimizing the influence of conscious processing*, a criterion often met through the use of various forms of subliminal/suboptimal stimulus presentation (e.g., short presentation durations, forward and/or backward masking, the use of continuous flash suppression).[1] It is standard practice to assess participants' conscious (un)awareness of suboptimal stimuli by conducting an (un)awareness check using an (un) awareness measure (Merikle & Cheesman, 1987). Yet the assessment of stimulus awareness is a difficult issue in masked priming studies that has no consensus. Indeed, there is no standard for what constitutes participants' conscious awareness versus unawareness, nor what measure(s) is (are) best to assess participant awareness. This looseness manifests in studies utilizing ostensible suboptimal presentations that largely fail to explicitly operationalize participant awareness. Is awareness established by the mere objective detection of something presented? Does awareness require only the recognition of what was presented? Or does awareness necessitate an explicable knowledge of what was presented? Imagine that people presented small sheets on which is printed a single number or letter are tasked with simply reporting the text (Sidis, 1898). Although presented at a distance rendering text nothing more than a reportedly blurry dot, participants were better than chance at forced-choice guessing both the category and content of the text. Here participants were aware that *something* was presented and could guess *what* it was but could not consciously *report* the content. Although some may consider these stimuli to have been presented outside consciousness, others may argue that mere detection or discrimination from alternatives constitutes awareness (Dehaene et al., 2006; Marcel, 1983).

Without a standard for what constitutes awareness, researchers must decide for themselves how to measure it. This has resulted in methodological issues across research utilizing suboptimal presentation since different awareness measures (e.g., detection tasks, alternative forced-recognition tasks, and verbal reports, among others) reflect distinct operationalizations of awareness, and researchers rarely detail or justify why a particular measure was chosen (van der Ploeg et al.,

1. Research employing quick stimulus presentations often uses the term *subliminal* to reference stimuli presented ostensibly below the threshold of conscious perception. The Latin root of the word, *limen*, means "threshold," so speaking of stimulus presentations as subthreshold (i.e., subconsciousness) has become the default language. Throughout this article I instead refer to degraded stimulus presentation meant to limit awareness as "suboptimal." The term *suboptimal* more accurately captures the continuum of awareness as it merely indicates that something has been presented at less than the highest quality at which it could be presented. In this case, suboptimal refers to the degradation of one's ability to consciously perceive and process a stimulus. As the current work will make clear, there are many levels below complete conscious perception at which a researcher may measure awareness. The term *subliminal* is therefore both inaccurate and misleading. The term *suboptimal* is both more reflective of the gradated nature of awareness and is accurately used when describing the design and findings from all awareness measures. I suggest researchers employing quick stimulus presentations will more clearly convey presentation paradigms by using the term *suboptimal* when describing study designs.

2017). Nonetheless, researchers have and continue to make claims of participant (un)awareness based on idiosyncratic operationalizations of "awareness" and unstandardized practices for testing participant awareness. This issue matters beyond purely methodological reasons; it also has implications for the fundamentals of empirical knowledge and theory. The data presented here, which clearly reflect this issue, are the result of various efforts measuring participant awareness of different classes (e.g., positive, neutral, negative, threatening) of objects across several studies examining physiological responses to stimuli presented outside awareness (March et al., 2022). Across studies I found that threat stimuli were processed preferentially. Yet although the trial structure (i.e., stimulus duration, masking parameters) remained nearly constant across studies and awareness measures, varying the discrimination threshold (i.e., the *type of awareness* needed for a correct response; Bevan, 1964) between tasks led to different conclusions about participant awareness. Had I relied on an awareness measure indicating participant unawareness, I may have theorized that the unique processing of threat was primarily the result of nonconscious processing. Alternatively, had I instead relied on an awareness measure indicating participant awareness, I may have theorized that the unique processing of threat requires some degree of conscious processing. The choice of awareness measure is therefore more than just a manipulation check but is an integral decision whose outcome can change the conclusions of entire programs of research.

This necessary consideration exemplifies two challenges common to research using suboptimal presentations. First, the use of various awareness measures between studies makes it difficult to interpret claims about suboptimality and participant unawareness within any given study. This means that beyond the conclusions of one program of research, the atheoretical choice of measures makes moving the field forward difficult as extensions of prior work might not be sensitive to *how* awareness was previously measured. Second, awareness is undefined (or, perhaps, undefinable) if it is a relative standard characterized within the confines of the measure used to test for it. That is, if the operationalization of awareness is directly linked to how one chooses to measure awareness (Reingold & Merikle, 1990), it blurs the distinction between measurement and manipulation and says nothing about participant awareness per se.

This is not a purely contemporary issue; many recent and past works have discussed the difficulty of measuring awareness and offered important perspectives and recommendations (e.g., Breitmeyer & Öğmen, 2006; Cheesman & Merikle, 1984; Eriksen, 1960; Fisk & Haase, 2005; Koster et al., 2020; Merikle, 1984; Michel, 2023; Newell & Shanks, 2014; Ramsøy & Overgaard, 2004; Rothkirch & Hesselmann, 2017; Snodgrass, et al., 2004; Stein & Peelen, 2021; Timmermans & Cleeremans, 2015; Wiens, 2005, 2006). Unfortunately, these recommendations have by and large been ignored, and measurement issues are continually overlooked in work employing suboptimal stimuli. As I demonstrate in the current article, ignoring these issues can create problems for research and researchers that rely on presenting stimuli beyond awareness. A consideration of best practices is therefore appropriate, timely, and necessary. The current work adds to an ongoing discussion of the methodology of the field by drawing attention to how operational definitions and tasks impact the results obtained from these experiments. I will not argue for an independent standard of unawareness. On the contrary, I argue that there is no objective standard for what constitutes "awareness" and, furthermore, that such a standard may not be possible, helpful, or even necessary to achieve most research (and researchers') goals. Indeed, it has been suggested that "defining and measuring awareness is conceptual" (Wiens, 2006, p. 118). In the current work, I argue that defining awareness depends critically upon the tasks chosen to document stimulus awareness. The current work does not focus on relating measures to conceptualizations of phenomenal or accessible perception. I also will not argue in favor of any class or type of awareness measure. Whereas prior work has argued in favor of objective versus subjective type measures based on discrete operationalizations of awareness (e.g., Wiens, 2005), I instead argue that more important than defining an elusive empirical standard for unawareness is considering for what purpose limiting awareness is being used. That is, consideration must be given to the *degree or type of awareness* that must remain below the threshold of consciousness for any given research question. Through careful consideration of this question, a researcher can identify an appropriate awareness measure. While I only briefly touch on theory of nonconscious processing, many in-depth discussions of the theoretical distinctions between levels of nonconscious and conscious processing are available (e.g., Snodgrass et al., 2004).

I begin by discussing the concept of awareness and then overview a selection of commonly used awareness measures. I then present my work testing awareness across three studies each using different oft-employed awareness measures that render different empirical conclusions across studies. Though I do not provide an exhaustive test of every awareness measure used in the literature, the ones I do employ reflect the basic structure of some of the most prevalent measures. I conclude by discussing issues of language in this domain and offer suggestions for choosing an awareness measure that reflects one's research goal. I contend that such goals, as the opening quotation highlights, likely constitute a middle ground between nil and total awareness.

ON MEASURING AWARENESS

Instead of viewing awareness as an all-or-none phenomenon (i.e., one is either totally aware or totally unaware), it may be seen as a gradual process spanning nil to complete awareness, between which lie gradients of partial awareness (Dehaene et al., 2006; Gelbard-Sagiv et al., 2016; Lähteenmäki et al., 2015; Mangan, 2001; Sandberg et al., 2010; Wiens, 2006). This implies that *(un)awareness* can refer to distinct levels along a continuum (i.e., states of awareness; Overgaard et al., 2006). For example, one may (a) be aware that a quickly shown stimulus was presented, (b) experience diffuse positive or negative affect in response to that stimulus, and/or be able to (c) match features of that suboptimal stimulus to a selection of alternatives, all without *full conscious perception* (i.e., the ability to identify the stimulus, unaided by information external to the perceiver) of the suboptimal stimulus.

Conscious perception (i.e., the ability to describe a stimulus, unaided) is itself another state along the continuum of awareness (Fisk & Haase, 2005). Increasing levels of awareness may be commensurate with the "amount" of processing, with, for example, feed-forward processing leading to low-level "preattentive" perception and iterative reentrant processing leading to higher-level "attentive" perception (Di Lollo, 2018). The idea here is that perception is a dynamic process involving feed-forward and feedback connections between lower and higher visual areas.

Awareness as a continuum necessarily implies that different awareness measures may tap different *states, types, or levels of awareness*. These states have been roughly divided into those constituting *subjective* versus *objective* awareness. A subjective state of awareness refers to the "level at which subjects claim not to be able to discriminate perceptual information at better than at a chance level," while an objective awareness is the "level at which perceptual information is actually discriminated at a chance level" (Cheesman & Merikle, 1984, p. 391). In terms of awareness measures, the subjective versus objective distinction refers to the content of the measurement. As Ramsøy and Overgaard (2004) describe, "subjective" refers to measures that tap idiosyncratic states where subjects are given less restrained conditions to report their experiences. Stein et al. (2021) state that "awareness measures can be either subjective (based on participant's report) or objective (based on perceptual performance)" (p. 1) and note that the subjective "approach is to simply ask participants to introspectively report their experience of a barely perceivable (e.g., masked) stimulus" (p. 2). Subjective awareness therefore requires a perceiver to, for example, classify the phenomenological experience or identify the stimulus *unprompted* (Lähteenmäki et al., 2015). Alternatively, objective awareness requires that perceivers discriminate a correct versus incorrect choice in, for example, some form of object detection (i.e., yes/no) or alternative forced-choice (AFC) task (Pessoa et al., 2005). In essence, the distinction between subjective versus objective measures is in their discrimination thresholds (i.e., the state of awareness needed for correct response) and the type of information measured.

A subjective state of awareness requires that the perceiver classify the subjective experience or identify the stimulus unprompted (Ramsøy & Overgaard, 2004; Stein et al., 2021; Wiens, 2006). In measures of subjective awareness, no information is provided, and the participant is asked what was presented or to describe their phenomenological experience verbally or using a scale (e.g., the Perceptual Awareness Scale) and/or their confidence in such reporting (Dienes & Perner, 2004; Dienes & Seth, 2010). The goal of such tasks is to gather precise introspective reports from participants that allow for indexing various levels of nonconsciousness while measuring subjective conscious experience (Overgaard et al., 2006). In some tasks, participants are asked to rate how visible a stimulus was on a continuous scale. The advantage of such measures is they are thought to index awareness of the stimulus and not simply awareness of one's awareness (Wierzchoń et al., 2014). In other tasks, participants express confidence in their awareness, and awareness is indirectly gleaned from wagers or ratings. From the subjective view, if the goal of suboptimal presentation is to prevent conscious (i.e., explicable) perception, anything less than a correct self-report can be characterized as unaware or a more granular categorization of levels of nonconsciousness may be possible (Overgaard & Sandberg, 2021). A focus on the subjective experience assumes that awareness is more a process of noticing than discriminating, and so an awareness measure ought to index what people notice instead of only what they can discriminate (Wiens, 2006; Wiens & Öhman, 2002). A critique of such measures suggests they may reflect participant-level biases to respond in a certain manner and may be conflated by the reporting itself (e.g., Goldiamond, 1958; Irvine, 2012). Response biasing can result from innate dispositional factors (e.g., affective responses associated with certain responses that increase/decrease their likelihood of occurrence) or features of the measure that encourage/discourage certain responses (Eriksen, 1960). Or, perhaps some individuals are more tuned to their introspective states and able to obtain objective experiences from subjective experiences (e.g., via neurophenomenology; Bitbol & Petitmengin, 2017), granting them insights others may not possess.

The alternative is to measure an objective state of awareness instead of utilizing a report (Tsuchiya et al., 2015). An objective state of awareness requires that the perceiver identify a stimulus, recognize it, match a visual feature of the stimulus, or perhaps intuit the stimulus category based on felt affect (Storbeck & Clore, 2008). Absent from objective tasks is information about subjective awareness prior to providing participants with disambiguating information (i.e., what they noticed). Consider a yes/no detection task that measures awareness as discriminating the presence (or absence) of a suboptimal stimulus. A subset of trials contains target stimuli and others contain either foils (e.g., scrambled images) or nothing. The participant's task is to simply indicate whether something was presented. A correct response results from detecting the mere occurrence of something. The objective state of awareness in this task is operationalized as something seen (e.g., change in luminance), not what was seen. Similarly, alternative forced-choice (AFC) tasks measure how well one can discriminate a suboptimally presented image among one or many foils. Some information is provided to the observer in the form of both the correct and alternative choices, and a correct response occurs either by seeing the quickly presented image, or merely matching a feature (e.g., an area of high contrast) of the suboptimal stimulus to a response option. Some AFC designs instead require that participants categorize the valence or affective content of the stimulus (e.g., positive vs. negative). For instance, one trial may present a masked fearful face that participants categorize as either "fearful" or "happy." Some information is again provided (i.e., the options of either fearful or happy), and a correct response occurs either by seeing the suboptimal stimulus, or by using nonconscious affective (i.e., gut) feelings or discriminating a unique feature (i.e., bright eye sclera) to guide response categorization of the face (Bar-Anan & Amzaleg-David, 2014). Although many now argue that subjective measures index conscious awareness better than objective tasks or verbal reports (Dienes & Seth, 2010; Koch & Preuschoff, 2007; Sandberg et al., 2010; Wierzchoń et al., 2012), others suggest the inverse, and indeed both classes of measures (and many types of measures) are still prevalent in work employing suboptimally presented stimuli (e.g., verbal/written report, forced-recognition task, threshold detection task, familiarity task, etc.; van der Ploeg et al., 2017)

ON ISSUES WITH MEASURING AWARENESS

Although the subjective/objective distinction is helpful for highlighting that different measures likely index different states of awareness, it does little to uncouple awareness from measurement. What constitutes awareness remains measurement contingent as a researcher's conclusions about awareness are directly linked to measurement (Dehaene et al., 2006). That is, were one to use a subjective standard, they may conclude that participants are quite unaware; but were they to use an objective standard, they may come to a different conclusion. This is the experience portrayed through the subsequently described data. My goal is to highlight the risk of defining awareness indirectly through measurement while disregarding distinctions between measures and, perhaps more importantly, the researcher's goal for using suboptimal stimuli. Treating all awareness measures as reflecting the same underlying phenomenon fails to consider how differences among measures reflect important information about the very processes under investigation. Indeed, at least for some tasks/assumptions, there is a quantitative relationship between measures, like detection and identification (MacMillan & Creelman, 1991).

The takeaway from the current work is that an awareness measure should reflect the conditions and characteristics of the awareness one wishes to measure. Using a one-size-fits-all approach to measuring awareness will likely fail to capture the relevant stimulus dimension by, for example, ignoring the unique operational conditions under which participants are undertaking a main task (i.e., under what state of awareness). Considering these differences will clarify the underlying process and result in a cleaner understanding of how different levels of awareness differentially impact the critical output and response while granting the researcher confidence in their manipulation. I present data from three studies measuring participant awareness using three distinct measures. These awareness measures were originally included in my research exploring individuals' physiological responses to suboptimally presented stimuli (March et al., 2022). That work involved three studies that each employed the same stimuli and nearly identical stimulus presentation parameters. I utilized the same stimuli in testing awareness by employing across three separate samples (1) an object detection task (ODT; i.e., yes/no) to assess if people can objectively discriminate the presence of a stimulus, (2) an alternative forced-choice (AFC) task to measure if people can objectively discriminate the presented stimulus out of a lineup, and (3) a subjective self-report (SSR) that assessed how well people could subjectively report what they had noticed. I begin by briefly describing each measure and then present data from all three measures simultaneously to enable the reader's comparison of participant awareness levels across the different measures. The trial totals, stimulus durations, and masking procedures in each study match those from their paired study in March et al. (2022), of which two studies included sandwich-masked stimuli (i.e., pre- and backward masked; indicated as SM) and one study included a prestimulus fixation and a backward mask (e.g., backward mask only; indicated as BM). Consequently, in each awareness measure, between-subjects conditions differed in their masking technique to assess whether awareness varied as a function of masking technique × measure.

The original motive for this work was to ensure that, on average, participants were unaware of the stimuli they were presented (i.e., at a chance level of accuracy). This could be done in two ways: (1) by measuring an overall level of accuracy on a specific awareness measure, agnostic to stimulus category (i.e., across all categories); or (2) by assessing whether awareness varied as a function of the stimulus type (e.g., positive, negative, neutral, or threatening). Indeed, the original work presented four different categories of stimuli, and the hypotheses of the main work (March et al., 2022) focused on testing for differences in physiological responses to different classes of quickly presented stimuli. Therefore, in each of the studies presented herein, I present awareness separately for each category. Lastly, previous work has indicated that individuals vary in their idiosyncratic ability to perceive quickly presented stimuli (Pessoa et al., 2005). Though focusing on each individual is beyond the scope of the current work, I descriptively explore between-subject differences in ability, and as I discuss in the General Discussion below, this is another area to consider when employing suboptimal stimuli. All data and analyses code are available in an online repository linked to this work at https://osf.io/cyx5s/?view_only=e769a9c8888446229359a92c8efc745c.

STUDIES TESTING AWARENESS

STIMULI

All studies presented stimuli employed by March et al. (2022) testing physiological responses to suboptimally presented stimuli (see Figure 1 and the Supplemental Material for all stimuli; all images are in the pubic domain).

Those stimulus sets were originally developed and piloted in earlier work by March et al. (2017) examining physiological responses to those same stimuli presented optimally. Categories were equated for luminance and red correction, which is necessary as systematic differences in category-level luminance, color, or other low-level visual attributes are often confounded with emotional categories (Lakens et al., 2013).

MEASURES AND METHODS

Yes/No in an Object Detection Task (ODT)

Participants. Participants completed an object detection task that was backward masked (i.e., ODT-BM, $n = 46$) or that was sandwich masked (i.e., ODT-SM, $n = 135$).

FIGURE 1. Examples of stimuli used in all awareness measures. See Supplemental Material for all stimuli. All images are in the public domain.

Study Apparatus and Procedure. All participants were seated in a cubicle ~75 cm from a 60 cm 144 hz high-speed monitor. Participants were told that we were interested in how well people detect images presented very quickly. They were informed that during the trial, "*Sometimes an image would be flashed very quickly, while other times no image would be presented*." Their task was to indicate whether an image was presented. Each trial in ODT-BM began with a fixation that was replaced by a 14 ms duration stimulus that was followed by a backward mask (colorful mosaic). Each trial in ODT-SM began with a 2 s pre-mask (colorful mosaic) that was replaced by a 21 ms duration stimulus (note the 7 ms duration difference between BM and SM; this was due hardware differences between computers but was constant within each condition and results in a more liberal test of awareness in the SM condition) that was followed by a backward mask (the same colorful mosaic rotated 90^o). The trial concluded with a prompt asking whether they thought an image was presented (see Figure 2). Participants used the z key to indicate an image was presented or the forward slash (/) key to indicate no image was presented. In ODT-BM, participants completed 128 trials, of which 64 (50%) contained a stimulus (16 negative, 16 neutral, 16 positive, and 16 threatening). In ODT-SM, participants completed 100 trials, of which 50 (50%) contained a stimulus (12 or 13 [varied randomly between subject] negative, 12 or 13 neutral, 12 or 13 positive, and 12 or 13 threatening, always adding to 50).

Statistical Analyses. In detection tasks, individuals often evidence a bias towards responding with either yes or no. Such bias makes interpreting proportions of

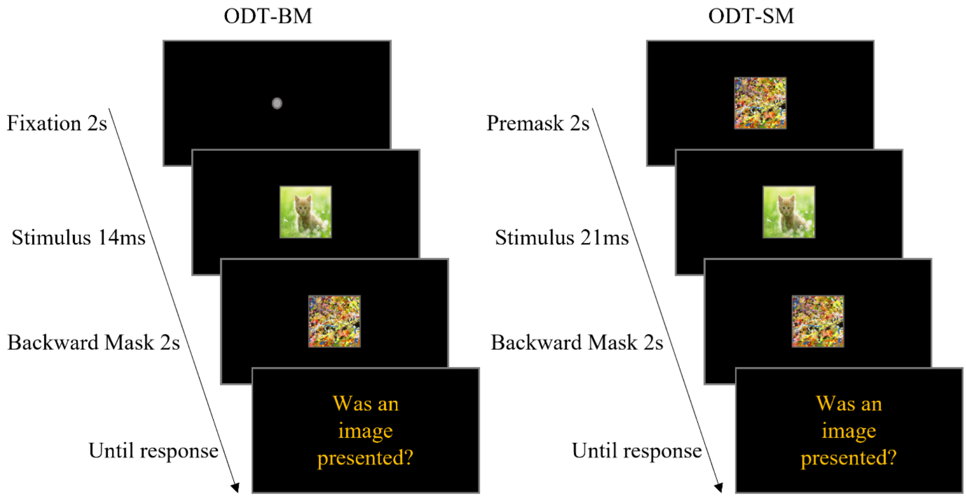


FIGURE 2. Diagram of the sequence of images presented on an ODT trial.

correct responses difficult. Therefore, I relied on signal detection theory (SDT; MacMillan & Creelman, 1991; Stanislaw & Todorov, 1999) to calculate each participant's accuracy in detecting quickly presented stimuli. SDT provides estimates of both sensitivity to the presence of a stimulus (d') and bias towards saying yes or no (c). I first calculated d', the difference between standardized (z score) correct yes responses when a stimulus was present (i.e., zH; hits) and incorrect yesses when no stimulus was present (i.e., zFA; false alarms; $d' = z\text{H} - z\text{FA}$). It is necessary to avoid the z transformation reaching infinity when hits or false alarm rates are perfect, as in when a subject adopts an extremely conservative or liberal criteria (i.e., responds yes or no on every trial). As is standard, values of 0% were recoded as $p = 1/N$ where N is the number of trials used in the calculation of p, and values of 100% were recoded as $p = (N - 1)/N$ (MacMillan and Creelman, 1991). The values for d' typically range from 0 to 4 and are interpreted as the difference between the distribution of probabilities of a hit in relation to a false alarm. I also calculated c, which is measured in standard deviation units, and is the bias toward saying yes (i.e., something was present) or no (i.e., nothing was present) on any given trial ($c = -0.5 \times [z\text{H} + z\text{FA}]$). When c is 0, there is no bias toward responding yes or no; a negative c indicates a liberal bias (i.e., more likely to say yes than no), while a positive c indicates a conservative bias (i.e., more likely to say no than yes). In yes/no tasks like the one used here, d' assumptions are often violated (e.g., unequal signal-to-noise standard deviations). In these cases, d' varies as a function of response bias c. Hence, I also calculated the nonparametric measure of sensitivity a', which is analogous to d' but accounts for the correlated nature of d' and c (Stanislaw & Todorov, 1999; cf. Zhang & Mueller, 2005). The value of a' typically ranges from .5 (signal indistinguishable from noise) to 1 (perfect performance).

Feature Discrimination in an Alternative Forced-Choice (AFC) Task

Participants. A separate cohort of participants completed an eight-alternative forced-choice task that was backward masked (i.e., 8-AFC-BM; $n = 46$) or a three-alternative forced-choice task that was sandwich masked (i.e., 3-AFC-SM; $n = 80$). (The number of forced-choice options was changed from eight equally divided among four categories in the BM study to three from only one category in the SM study to address a design flaw in 8-AFC-BM. In that task, participants showed a response bias toward selecting only certain categories [i.e., neutral and positive], inflating accuracy for these categories, and seemed to actively avoid selecting other categories [i.e., negative and threatening], diminishing accuracy for those categories. This results in a confound and precludes comparing results from the 8- and 3-AFC tasks. Further details on this are given below.)

Study Apparatus and Procedure. Participants were seated in a cubicle ~75 cm from a 60 cm 144 hz high-speed monitor. In both studies, participants were told that "*We are interested in how well people can identify images that are presented very quickly*." They were informed that during the trial, an image would be flashed very quickly. Their task was to choose which image was presented from a selection of images. Each trial in the 8-AFC-BM began with a fixation that was replaced by a 14 ms duration stimulus that was backward masked by a 3000–5000 ms mosaic (the same colorful mosaic rotated 90^o). Each trial in the 3-AFC-SM began with a 2 s pre-mask (colorful mosaic) that was replaced by a 21 ms duration stimulus that was backward masked by a 3000–5000 ms mosaic (the same colorful mosaic rotated 90^o). The trial concluded with a screen presenting a selection of options from which they indicated those previously presented. In the 8-AFC-BM, participants were presented with a matrix of eight choices of which two images each were negative, neutral, positive, or threatening. In the 3-AFC-SM, participants were presented with a matrix of three choices in which each image was from the same category as the suboptimal image (see Figure 3).

Participants used the keyboard number pad to indicate which image was presented. In the 8-AFC-BM, participants completed 64 trials, of which 16 each were negative, neutral, positive, or threatening. In the 3-AFC-SM, participants completed 36 trials, of which nine each were negative, neutral, positive, or threatening.

Conscious Perception in Subjective Self-Report (SSR)

Participants. A separate cohort of participants completed a subjective self-report task that was backward masked (i.e., SSR BM, $n = 43$) or that was sandwich masked (i.e., SSR-SM, $n = 50$).

Study Apparatus and Procedure. Participants were seated in a cubicle ~75 cm from a 60 cm 144 hz high-speed monitor. In both studies, participants were told that "*We are interested in presenting images so quickly that they cannot be seen*." They were informed that during each trial, an image would be flashed very quickly. To

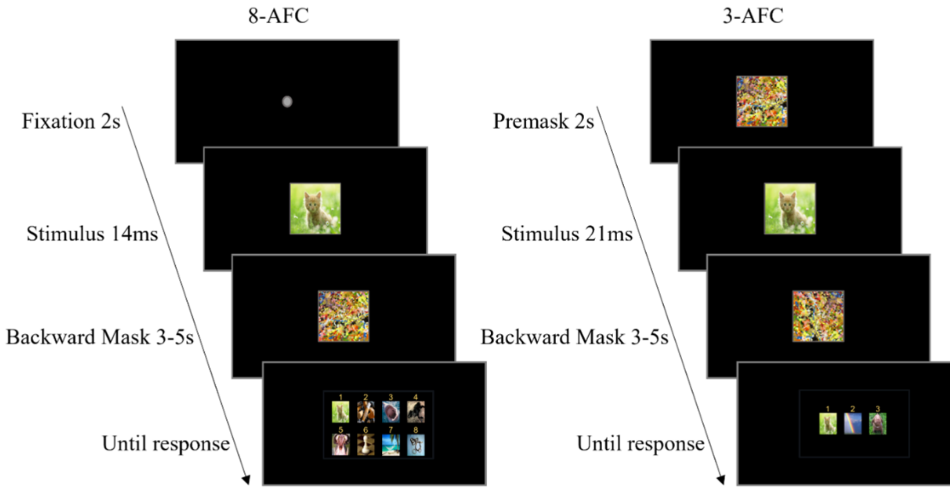


FIGURE 3. Diagram of the sequence of images presented on an AFC trial.

motivate attention and accuracy, participants were told they would receive $1 for each correct answer such that they could earn up to $64 in the SSR-BM or $100 in the SSR-SM. There was no penalty for incorrect answers. To rule out laziness and quickly clicking to the end, we required participants to type a response on each trial (a brief description or "do not know"). It was expected that this would encourage them to guess even if they were uncertain. Each trial in the SSR-SM began with a 2000 ms centrally located pre-mask (colorful mosaic) that was replaced by a 25 ms duration stimulus that was backward masked by a 3000–5000 ms mosaic (the same colorful mosaic rotated 90°). Each trial in the SSR-BM began with a fixation that was replaced by a 25 ms duration stimulus that was backward masked by a 3000–5000 ms mosaic (the same colorful mosaic rotated 90°). The trial concluded with a text box in which they indicated what was presented or typed "do not know" (see Figure 4). In the SSR-BM, participants completed 64 trials, of which 16 each were negative, neutral, positive, or threatening. In the SSR-SM, participants completed 100 trials, of which 25 each were negative, neutral, positive, or threatening.

METHODOLOGICAL VARIABILITY

Before turning to the results, I emphasize two points. First, there is methodological variability between the tasks employed in the current work. The previous sections detailed the presentation paradigms employed in each task, and to aid the reader in interpreting the subsequent results, the current section quickly details their distinctions. Across paradigms, display parameters are not perfectly equivalent both between paradigm and within paradigm between masking conditions. Recall that these awareness measures were originally conducted to reflect several

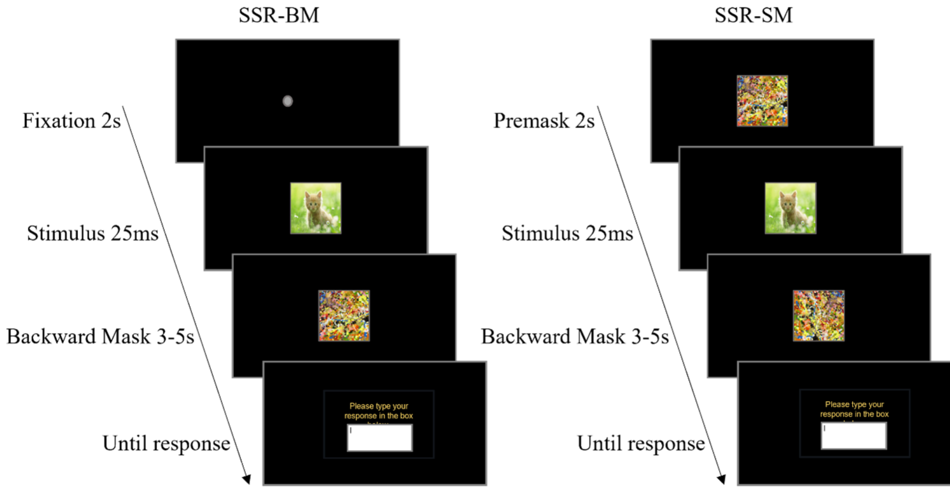


FIGURE 4. Diagram of the sequence of images presented on an SSR trial.

studies measuring physiological responses to suboptimally presented stimuli (March et al., 2022). The studies in that package comprised designs with distinct presentation parameters, and subsequent tests of awareness were meant to reflect the breadth of presentation paradigms from the original studies. That inconsistency manifests in the current package as minor variability. Specifically, (a) the duration of the backward masks varies from 2 to 5 s between studies and (b) the prime display times vary across experiments: 14 ms (detection backward masked, identification backward masked), 21 ms (detection sandwich masked, identification backward masked), or 25 ms (subjective self-report backward or sandwich masked). Although this incurs a limitation for making a consistent comparison across the three different tasks, that formal comparison is not a focal point of the research. Further, I do not expect that the noted variability significantly impacted the comparisons within tasks. I note it here for openness and transparency.

Second, different expressions of sensitivity and statistical analyses are used for detection versus identification versus self-report. The conventional approach was used to analyze data from each paradigm, consistent with what is commonly found in the literature. Specifically, the detection analysis is based on d', a' (unbiased), and c from signal detection theory with analysis of variance (ANOVA) tests. Identification is based on the proportion correct (potentially biased) and ANOVAs. Self-report was based on proportion correct and multilevel logistic regression. These statistical inconsistencies prevent a straightforward comparison across the three experiment task types. The different statistics for different methods (detection, identification, self-report) could be seen as a methodology flaw. Instead, I consider this as further support for my major argument point. Different measures lead to different statistics that are traditional for each task, which can then lead to different conclusions about awareness. So, not only could one come to different conclusions depending on the measure one uses, but that choice also necessarily entails the analytic strategy often associated with that measure. It then becomes unfeasible to compare statistics or results across tasks because of the inconsistency between approaches. Discussed in depth in other work (e.g., Schmidt & Vorberg, 2006) is the idea that developing a consistent metric (e.g., d' or some other effect size) across all types of awareness tasks would serve to provide a more direct comparison between studies employing awareness measures. Unfortunately, a straightforward implementation of a uniform metric across the many varying designs of awareness measures has so far eluded the field.

RESULTS

OBJECTIVE DETECTION TASK (ODT)

As described above, signal detection metric d' describes sensitivity to the presence of a stimulus c and describes one's bias towards a certain response; in this case, a bias towards responding whether an object was present or no object was present. The metric a' is analogous to d' but accounts for the correlated nature of d' and c. See Table 1 for raw percentage correct and other metrics. As c is a direct derivation of d', inferential results are identical and not reported. On average, c was positive, indicating participants' conservative bias toward reporting no stimulus was presented.

Backward Masked Stimuli

d′. Awareness of all classes of stimuli was greater than chance (i.e., d' greater than 0; all *t*s > 5.26, all *p*s < .0001). See Table 1 for all response metrics. A repeated measures ANOVA revealed a main effect of stimulus type, $F(3, 43) = 8.34$, $p = .0002$, indicating that awareness varied as a function of the class of stimulus. Post hoc pairwise tests indicated that participants were better able to detect the presence of threatening than positive, $F(1, 45) = 4.20$, $p = .0462$, $d = 0.30$, and negative, $F(1, 45) = 23.63$, $p < .0001$, $d = 0.71$, but not neutral stimuli, $F(1, 45) = 0.14$, $p = .7079$, $d = 0.06$; better able to detect neutral than positive, $F(1, 45) = 4.83$, $p = .0332$, $d = 0.32$, and negative stimuli, $F(1, 45) = 17.41$, $p = .0001$, $d = 0.62$; and better able to detect positive than negative stimuli, $F(1, 45) = 4.88$, $p = .0323$, $d = 0.33$.

a′. Awareness of all classes of stimuli was greater than chance (i.e., a' greater than 0.5; all *t*s > 3.97, all *p*s < .0004). There was a main effect of stimulus type, $F(3, 43) = 8.64$, $p = .0001$. Post hoc pairwise tests indicated that participants were better able to detect the presence of threatening than positive, $F(1, 46) = 8.67$, $p = .0051$, $d = 0.43$, and negative, $F(1, 46) = 26.36$, $p < .0001$, $d = 0.76$, but not neutral stimuli, $F(1, 46) = 2.51$, $p = .1202$, $d = 0.23$; better able to detect neutral than negative, $F(1, 46) = 8.87$, $p = .0047$, $d = 0.44$, but not positive stimuli, $F(1, 46) = 1.63$, $p = .2086$,

TABLE 1. Response Metrics for ODT-BM and ODT-SM

	Backward Mask Only				Sandwich Mask			
Stimulus	**% correct**	**d′**	**c**	**a′**	**% correct**	**d′**	**c**	**a′**
Negative	35	0.92	.927	.637	35	.534	.956	.569
Neutral	47	1.27	.758	.708	41	.692	.877	.618
Positive	40	1.08	.850	.670	35	.526	.960	.577
Threatening	46	1.24	.770	.746	41	.669	.888	.607

Note. d' is a measure of sensitivity; c is a measure of bias; a' is a parametric and less biased measure of sensitivity.

$d = 0.23$; and no better able to detect positive than negative stimuli, $F(1, 46) = 2.38$, $p = .1297$, $d = 0.18$.

Sandwich-Masked Stimuli

d'. Awareness of all classes of stimuli was greater than chance (i.e., d' greater than 0; all ts > 8.60, all ps < .0001). A repeated-measures ANOVA revealed a main effect of stimulus type, $F(3, 132) = 9.64$, $p < .0001$. Post hoc pairwise tests indicated that participants were better able to detect the presence of threatening than positive, $F(1, 134) = 17.57$, $p < .0001$, $d = 0.36$, and negative, $F(1, 141) = 15.91$, $p < .0001$, $d = 0.34$, but not neutral stimuli, $F(1, 141) = 0.51$, $p = .4752$, $d = 0.06$; better able to detect neutral than positive, $F(1, 134) = 20.18$, $p < .0001$, $d = 0.39$, and negative stimuli, $F(1, 134) = 18.01$, $p < .0001$, $d = 0.37$; and no better able to detect positive than negative stimuli, $F(1, 134) = 0.07$, $p = .7954$, $d = 0.02$.

a'. Awareness of all classes of stimuli was greater than chance (i.e., a' greater than 0.5; all ts > 4.11, all ps < .0001). A repeated-measures ANOVA revealed a main effect of stimulus type, $F(3, 132) = 6.43$, $p = .0004$. Post hoc pairwise tests indicated that participants were better able to detect the presence of threatening than positive, $F(1, 134) = 6.21$, $p = .0139$, $d = 0.21$, and negative, $F(1, 134) = 9.22$, $p = .0029$, $d = 0.26$, but not neutral stimuli, $F(1, 134) = 0.83$, $p = .3638$, $d = 0.08$; better able to detect neutral than positive, $F(1, 134) = 13.11$, $p = .0004$, $d = 0.31$, and negative stimuli, $F(1, 134) = 15.06$, $p = .0002$, $d = 0.33$; and no better able to detect positive than negative stimuli, $F(1, 134) = 0.64$, $p = .4267$, $d = 0.07$.

Idiosyncratic Abilities

Recall that chance performance in the a' metric equals 0.5 with 1 equaling perfect performance. As seen in Figure 5, there is indeed a wide distribution of performance ability with different individuals across studies ranging from below chance to nearly perfect performance. The negative skew of performance in backward

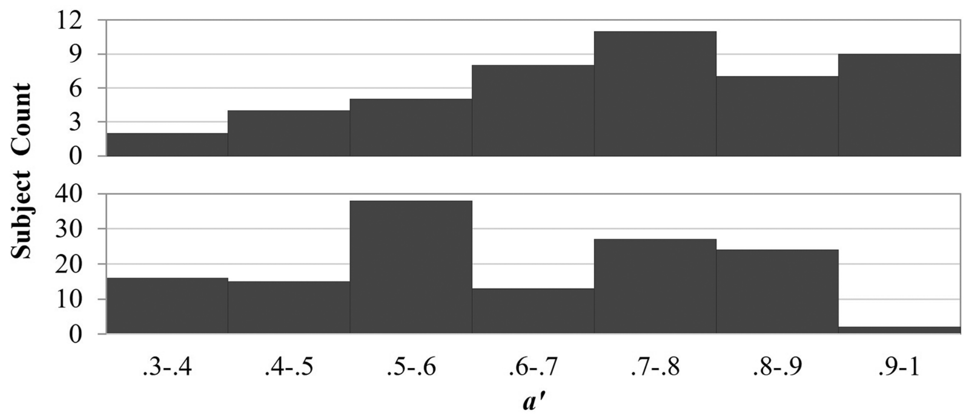


FIGURE 5. Histographic representation of *a′* in the object detection task for backward-mask (top panel) and sandwich-masked (lower panel) presentations.

mask only (upper panel) versus the positive skew of sandwich-masked (lower panel) presentation indicates that a larger proportion of people performed above chance when stimuli were merely backward masked than sandwich masked. This is reflected in the mean *a′* of the backward-masked stimuli (.72), which is below its median (.77), while, conversely, the mean *a′* of the sandwich-masked stimuli (.61) is higher than its median (.57).

Discussion

Participants displayed an above-chance ability to determine the presence of a suboptimally presented stimulus in an ODT. Some of this is likely due a consistent and conservative bias toward saying that nothing was presented, resulting in fewer false alarms than misses as evidenced by positive *c* values close to 1. This tendency was even more pronounced when stimuli were sandwich masked versus only backward masked (i.e., larger *c* values), indicating that participants had a harder time determining the presence of a stimulus during sandwich versus only backward masking. This interpretation is supported by descriptively smaller *d′* and *a′* (i.e., sensitivity) values when stimuli were sandwich masked versus backward masked.

There were also category level differences in awareness, with participants better able to determine the presence of a neutral/threatening than positive/negative stimuli. This may indicate that in an ODT, the mind is better able to perceive features more inherent in or sense the presence of certain types of stimuli than others (see March et al., 2022 for a discussion of this possibility for threat stimuli). Importantly, this does not indicate that people could identify the stimuli. Instead, detection tasks may be performed above chance level on the basis of simple luminance differences even when participants cannot identify what they have seen.

Secondarily but relevant to the aim of the current work, these results also indicate that awareness was lower under sandwich- versus backward-masked conditions even though there was bias toward responding with a no versus yes in both conditions. This pattern emerged despite the additional 7 ms stimulus-onset asynchrony (SOA) in the sandwich-masked condition, which one might expect would make stimuli more accessible. Sandwich masking appears to be a clearly superior form of masking (see Breitmeyer, 2015, for a review).

ALTERNATIVE FORCED-CHOICE (AFC) TASK

In both versions, the numerical position of the correct response option was random. In the 8-AFC-BM, the correct choice was equally distributed across the eight options, and in the 3-AFC-SM, across the three options. Given this design, no position response bias was expected (i.e., people were no more likely to choose response 1 than response 3, all else being equal) and so in lieu of signal detection, I used the traditional approach to scoring AFC tasks and calculated the more easily interpretable probability of a correct response for each trial type (DeCarlo, 2012).

Backward-Masked Stimuli

At first glance it appears that participants in the 8-AFC-BM task more accurately identified the suboptimally presented stimulus when it was neutral, least accurately when it was negative, and somewhere between for threatening and positive. Yet, this pattern appears to be driven by a design flaw—a flaw still seen in recent work (e.g., Ruiz-Padial & Vila, 2007) and discussed further in the General Discussion below. By examining Table 2, one can see that participants chose the neutral and positive options more frequently, regardless of the suboptimal stimulus category. One participant noted in their post-session feedback, "I couldn't tell what any of the images were, I just picked the pleasant ones." Another said "I . . . didn't want to press on the more violent images." Yet another indicated that "the images were disturbing. I did not like looking at the disgusting pictures." Although there were only 736 trials containing a neutral or positive stimulus, respectively (a total of 1,472 trials), a neutral selection was chosen 994 times and a positive selection was chosen 940 times (1,934 total, or ~66% of all responses). This is in contrast to the threatening selections, which were chosen 640 times while the negative selections were chosen 401 times (1,041 total, or ~34% of all responses).

Participants seemingly avoided choosing the negative or threatening stimuli. This resulted in artificially inflated accuracy especially for the neutral but also the positive categories, and suppressed accuracy for particularly the negative but also the threat categories. Although it does appear as though participants were the most accurate for neutral stimuli, the design confound makes it difficult to draw conclusions from these data. This type of unexpected response bias has design implications for ACF designs. For this reason, I altered the selection process in the 3-AFC-SM task such that all options were from the same category. That is, people could no longer avoid certain categories.

TABLE 2. Percent of Responses for Each Stimulus Category by Chosen Category in the 8-AFC-BM

	Chosen Category			
Stimulus Category	**Negative**	**Neutral**	**Positive**	**Threat**
Negative	11.6% + 5.6%	30.3%	35.2%	17.4%
Neutral	11.0%	37.5% + 11.8%	25.7%	14.0%
Positive	13.6%	31.4%	19.0% + 16.2%	19.8%
Threat	12.8%	24.0%	31.7%	21.0% + 10.5%
Average	13.7%	33.8%	32%	20.7%

Note. There were 736 trials within each stimulus category. In an eight-choice task, there were two choices from each category. One choice for any given stimulus was necessarily of the same category but of the incorrect stimulus. To grant the reader clarity into not only correct choices, but also category choices (i.e., category bias), the values surrounding the "+" in the diagonal cells represent the percentage of selection of the correct stimulus + the incorrect stimulus but from the correct category, respectively.

Sandwich-Masked Stimuli

Across stimulus types, participants were above chance (i.e., 3 of 9) at correctly choosing the presented stimulus out of a three-image lineup (all *t*s > 7.15, all *p*s <.0001). A repeated-measures ANOVA revealed a main effect of stimulus type, $F(3, 77) = 9.18$, $p < .0001$. Post hoc pairwise tests indicated that participants were less accurate when the stimulus was threatening ($M = 4.15$; out of 9) than neutral ($M = 5.31$), $F(1, 79) = 21.96$, $p < .0001$, $d = 0.52$, or positive ($M = 4.89$), $F(1, 79) = 10.90$, $p = .0014$, $d = 0.37$, but equally accurate when the stimulus was threatening versus negative ($M = 4.40$), $F(1, 79) = 1.16$, $p = .2851$, $d = 0.12$. Participants were also more accurate when the stimulus was neutral than positive, $F(1, 79) = 3.55$, $p = .0632$, $d = 0.21$, or negative, $F(1, 79) = 16.32$, $p = .0001$, $d = 0.45$, and more accurate when the stimulus was positive than negative, $F(1, 79) = 5.18$, $p = .0256$, $d = 0.25$. See Figure 6.

Idiosyncratic Abilities

Participants varied widely in their abilities to perform the 3-AFC-SM task. As can be seen in Figure 7, accuracy ranged from a total of 11 out of 36 (31%) to 26 out of 36 (72%) correct. The mean correct (18.75) is almost exactly at the median (19), indicating an unbiased and symmetric distribution of ability.

Discussion

Although the flawed design of the 8-AFC-BM task precluded making inferences about accuracy, it does imply that voluntary category response bias is an important consideration when designing forced-choice tasks. That is, the opportunity to avoid looking at or choosing certain classes of stimuli (e.g., gross, threatening, mutilation) can lead to the belief of a spuriously lowered awareness of these categories and, commensurately, artificially inflated awareness of others. This

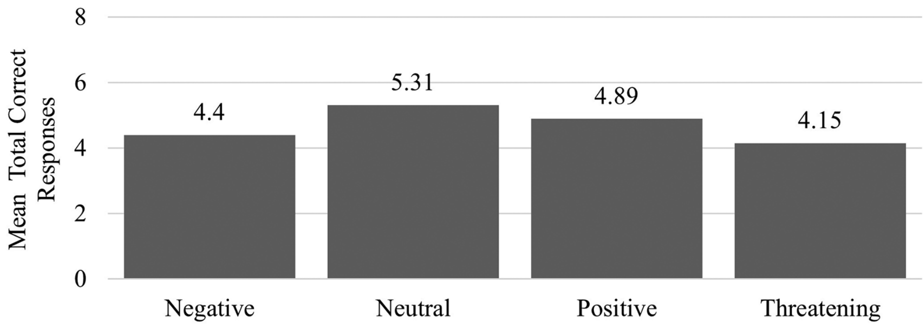


FIGURE 6. Mean accuracy in the 3-AFC-SM task as a function of stimulus category.

consideration may seem intuitive, but this type of response bias is still overlooked in recent research (including, obviously, my own) and is discussed further in the General Discussion below.

The results from the 3-AFC-SM task implied that people were most accurate when choosing the selection that matched suboptimally presented neutral stimuli and least accurate when matching the threatening stimuli. Accuracy to positive and negative stimuli fell in between. The inaccuracy to threat contrasts with the results of the ODT. This may be because awareness in an AFC versus ODT task is more the result of feature matching. Consider one participant's statement, "I could not really see the images, but it was almost like I could see outlines for a split second after the image had passed." It is well known that certain factors can affect the efficacy of masking, such as unique physical features like color, brightness, contrast, and resolution, features that may be more or less prevalent among different classes of stimuli. The results of the 3-AFC-SM task imply that the mind

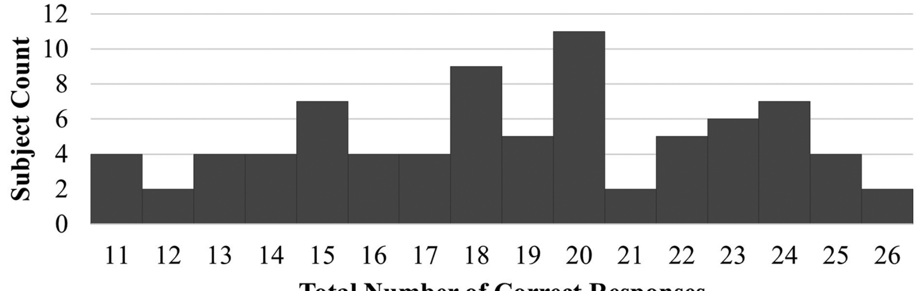


FIGURE 7. Histographic representation of correct response totals by each subject on the 3-AFC-SM task (out of 36 trials); chance is 12/36 correct.

may better perceive features more prevalent in certain stimuli—perhaps among these stimuli, those frequently present within the neutral (and perhaps, positive) set. Meaning, like the ODT task, these results do not necessarily speak to whether people were consciously aware of what they saw, but only that they were able to discriminate certain features.

SUBJECTIVE SELF-REPORT TASK

Two assistants blind to the purpose of the study coded responses for accuracy with instructions to use a liberal standard such that responses needed to be generally accurate but not exact. For example, it would be correct to respond "bug" to a cockroach, "bowl" or "cylinder" to a saucepan, or "burglar," "man in a mask," or "gunman" to a masked man holding a gun. The assistants agreed on 7,745 of 7,752 responses (99.91%), and I rectified the seven disagreements.

To test for differences in accuracy across the trial types, I conducted a multilevel logistic regression with a binary distribution using PROC GLIMMIX of SAS with a random intercept to control for the nested trial-by-trial responses within participants and a random intercept for each stimulus to account for stimulus-level variation (Judd et al., 2012).

Backward-Masked Stimuli

Participants were able to correctly identify 158 stimuli out of 2,752 total trials (5.74%). Across stimulus types, people were more likely to be incorrect than correct (all ts > 29.99, all ps < .0001; Eriksen, 1960; Goldiamond, 1958). In order of incorrectness, participants were 1105.84 times more likely to be incorrect than correct on negative trials, $b = 7.01$, $t(2646) = 8.57$, $p < .0001$, 147.70 times more likely to be incorrect on threat trials, $b = 4.99$, $t(2646) = 8.48$, $p < .0001$, 80.33 times more likely to be incorrect on positive trials, $b = 4.39$, $t(2646) = 7.81$, $p < .0001$, and 28.11 times more likely to be incorrect on neutral trials, $b = 7.09$, $t(2646) = 6.33$, $p < .0001$. See Table 3 for response totals.

Though people were more likely to be incorrect than correct across stimulus types, there was a main effect of stimulus type, $F(3, 2646) = 6.42$, $p = .0003$, indicating that levels of incorrectness varied by category. Post hoc pairwise tests indicated that participants were more often incorrect when the stimulus was negative ($M = 0.16$; out of 16) than neutral ($M = 1.98$), $t(2646) = 4.24$, $p < .0001$, positive ($M = 0.93$), $t(2646) = 2.91$, $p = .0036$, or threatening ($M = 0.60$), $t(2646) = 2.18$, $p = .0297$; more often incorrect when the stimulus was threatening than neutral, $t(2646) = 2.51$, $p = .0121$, but not positive, $t(2646) = .88$, $p = .3814$; and equally likely to be incorrect when the stimulus was positive or neutral, $t(2646) = 1.67$, $p = .0958$.

Sandwich-Masked Stimuli

People were able to correctly identify 34 stimuli out of 5,000 total trials (.68%). Across stimulus types, people were more likely to be incorrect than correct (all

TABLE 3. Responses for SST-BM and SST-SM

	Backward Mask Only			Sandwich Mask		
Stimulus	**Don't Know**	**Incorrect Guess**	**Correct Response**	**Don't Know**	**Incorrect Guess**	**Correct Response**
Negative	503	178	7 (1%)	999	248	3 (0.2%)
Neutral	327	276	85 (12%)	780	449	21 (1.7%)
Positive	471	177	40 (6%)	997	247	6 (0.5%)
Threat	412	250	26 (4%)	951	295	4 (0.3%)

Note. Each row in SST-BM totals to 688 trials; each row in SST-SM totals to 1,250 trials. Correct Response columns contain the correct tally raw and as a percentage of total trials.

*t*s > 132.87, all *p*s < .0001). In order of incorrectness, participants responded incorrectly most often on negative trials, where they were 1180.24 times more likely to be incorrect than correct, $b = 7.07$, $t(4853) = 9.18$, $p < .0001$, followed by threat trials, where they were 936.20 times more likely to be incorrect, $b = 6.84$, $t(4853) = 9.29$, $p < .0001$, positive trials, where they were 585.64 times more likely to be incorrect, $b = 6.37$, $t(4853) = 9.96$, $p < .0001$, and neutral trials, where they were 175.17 times more likely to be incorrect, $b = 5.17$, $t(4853) = 9.81$, $p < .0001$. See Table 3 for response totals.

Although people were again more likely to be incorrect than correct across stimulus types, there was a main effect of stimulus type, $F(3, 4853) = 4.11$, $p = .0064$, indicating that levels of incorrectness varied by category. Post hoc pairwise tests indicated that participants were more often incorrect when the stimulus was negative ($M = 0.06$; out of 25) than neutral ($M = 0.42$), $t(4853) = 2.68$, $p = .0075$, but not positive ($M = 0.12$), $t(4853) = 0.90$, $p = .3686$, or threatening ($M = 0.08$), $t(4853) = 0.32$, $p = .7502$; more often incorrect when the stimulus was threatening than neutral, $t(4853) = 2.53$, $p = .0115$, but not positive, $t(4853) = .60$, $p = .5480$; and more likely to be incorrect when the stimulus was positive than neutral, $t(4853) = 2.10$, $p = .0356$.

Idiosyncratic Abilities

Awareness varied because of inter-individual differences in intrinsic ability to perceive quickly presented stimuli. As can be seen in Figure 8, the distributions of performance in both the SST-BM and SST-SM were positively skewed, though skewness is more extreme in the SST-BM than in the SST-SM task. The SST-BM had an overall correct mean of 3.67/64, almost twice the median of 2, and as can be seen a few people disproportionately affected the mean relative to the sample, and correct responses were unequally distributed among the participants. Indeed, in the SST-BM, 30 of the 85 total correct neutral responses (35%) were given by only 4 out of 43 total (9%) participants. Conversely, the mean correct tally in the SST-SM was .68/100 with a median of 0. Response accuracy was much more closely clustered in the SST-SM versus the SST-BM, likely reflecting the increased suppressing effect of the sandwich mask.

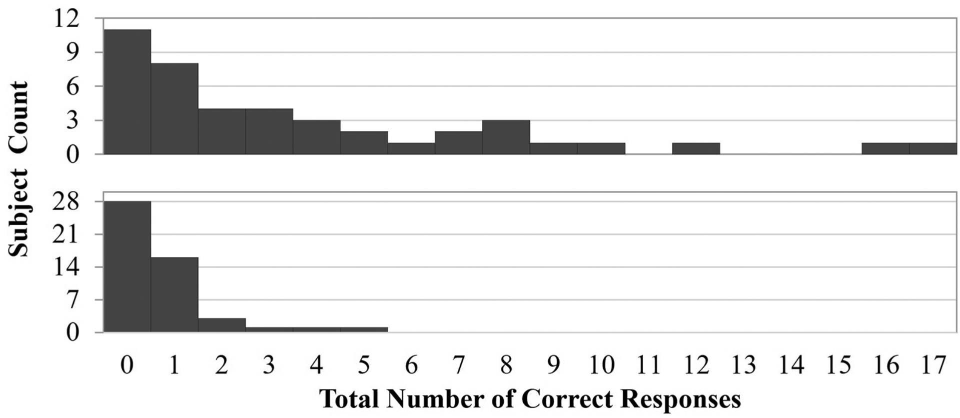


FIGURE 8. Histographic representation of correct response totals by each subject for the SST-BM (top panel; out of 64 trials) and the SST-SM (lower panel; out of 100 trials).

Discussion

Despite the monetary incentive to actively attend and accurately describe the stimuli, participants were remarkably unable to do so. Across both the SST-BM and SST-SM tasks, people were better able to identify the suboptimal stimulus if it was neutral than threatening, positive, or negative. Participants in the SST-BM task averaged less than 2 correct per 16 presented from each category and in the SST-SM task averaged less than .5 correct per 25 presented from each category. The presence of a sandwich versus backward mask made identifying the object more difficult across all categories of stimuli. Sandwich versus only backward masking appears to be a superior technique for rendering stimuli suboptimal. Yet as in the previous two studies, sensitivity was highest when identifying neutral stimuli.

GENERAL DISCUSSION

Research on how the mind processes or responds to information outside awareness relies on suboptimal presentations to present stimuli below conscious perception. Accounting for participant awareness is an important consideration. Unfortunately, awareness is a broad term describing many levels of perception. This means that awareness varies depending on how it is measured. I presented work showing how using different awareness measures (i.e., different operationalizations of awareness) can lead one to different conclusions about participant awareness. Across three tests using the same stimuli, masking techniques, and nearly the same ultrafast stimulus duration, I measured whether participants could: (1) indicate the presence versus absence of a stimulus in an object detection task (ODT), (2) match the stimulus to a selection of options in an alternative forced-choice (AFC) task, and (3) describe in their own words what was presented in a subjective self-report (SSR) task. Three methods seemingly point to different top-level conclusions about awareness, though they sometimes converge about category-level awareness. Mean performance was generally above chance for ODT and AFC tasks but suffered greatly when awareness was measured at the level of conscious perception by the SSR task. If I (or any researcher) were to rely on the current ODT or AFC task, I might conclude that participants were quite aware. Alternatively, were I (or any researcher) to rely on the current SSR task, I might conclude that participants were incredibly unaware. Yet, as I subsequently discuss, even within each measure, the question remains as to what level of unawareness is sufficient. Were my goal to reduce awareness to 0%, even the most conservative test of awareness used here (i.e., the SSR-SM) may have indicated participants were experiencing too high a level of awareness. The presence of a pre- and backward mask versus only a backward mask led to decreased awareness across tasks where comparison was possible. This implies that sandwich-masked designs may be more effective at inhibiting awareness, an important consideration for future work using suboptimal stimuli (Breitmeyer, 2015).

During the remainder of the General Discussion, I consider how researchers can navigate the tricky waters of measuring awareness and potential issues with conservative decision bias. I then turn to the discussion of other design issues leading to inconsistency in awareness research. I end with a discussion of how certain features or low-level visual attributes may render certain classes of stimuli unequally difficult to mask, making it all the more important that awareness is checked at the level of the category, and not study-wide (i.e., collapsed across critical categories).

ON MEASURING AWARENESS

Knowledge can come in many forms that may reflect unique “states” of awareness. The predominant awareness measures either index one’s conscious percept of a stimulus or gauge one’s ability to indicate the presence or match the features of a suboptimal stimulus to a selection of choices (other tasks have also been deployed). Simple knowledge that a stimulus was presented (i.e., in an ODT task), the ability to describe a stimulus’s affective content or match the presented stimulus to presented alternatives (i.e., in an AFC/IFC task), and/or write or verbally provide a description of the stimulus (i.e., in an SSR task) can all be considered states of awareness. Each is a type of knowledge, but they are not qualitatively equal in their possible influence on stimulus processing. Consider a skin-conductance study where participants are told to simply watch a seemingly nonsensical dynamic stream of mask images and are told nothing about the presence of suppressed stimuli. Were the stimuli to remain completely below even the lowest level of awareness (i.e., as measured by a simple yes/no ODT), one could conclude that the subsequent influence of the suboptimal stimuli was uninfluenced by either an orienting response or conscious processing. Yet were participants to become aware of the presence of a suboptimal stimulus—even if they remained unable to identify or spot the stimulus—the increased attention garnered by the stimuli may elicit an orienting response that would register as a skin-conductance response (SCR), potentially convolving any SCR elicited by the suboptimal stimuli. The choice of data-processing technique (i.e., to justify deconvolving the SCR) may depend on the researcher's ability to determine a certain degree of participant awareness. Or consider my earlier attempt to measure facial electromyography (fEMG) induced by startle-eyeblink where participants viewed suboptimally presented images and occasionally heard a loud noise (March et al., 2022). On each trial, they indicated the presence or not of a subliminal stimulus (an online ODT task; in reality, a stimulus was present on every trial). The mere awareness of suboptimally presented stimuli encouraged participants to squint, attempting to improve their perception, which effectively inhibited the startle-induced eyeblink responses.

The potential for the presence of a confounding implicit or explicit process influencing subsequent measurement increases commensurately with higher levels of awareness. Imagine a snake-phobic person viewing suboptimally presented snake and non-snake stimuli. A fully suppressed image would evoke only those processes capable of perceiving and responding to stimuli presented in such a degraded manner. If these are the processes of interest to the researcher, ensuring no higher level of awareness is paramount to showing strong evidence for the influence of only nonconscious processes. Even perceiving the presence of a sinusoidal shape absent explicable knowledge of a snake may activate both critical nonconscious and confounding conscious processes, precluding making strong claims about the influence of purely nonconscious processes. Yet were one interested in the influence of nonconscious and preconscious processes (i.e., aware but unable to identify), such awareness may be completely in line with their research goals.

These cases are only a narrow example of why this issue matters beyond purely methodological reasons (e.g., beyond helping researchers in this space design studies). Consider a phenomenon believed to occur automatically and outside awareness, for example, first impressions and thin slices of personality (Ambady et al., 2000; Willis & Todorov, 2006). Here it is theorized that people develop first impressions and can be accurate about personality traits after only 100 ms of exposure. Influential research has argued that the cues people use are "so subtle that they are neither encoded nor decoded at an intentional, conscious level of awareness" (Ambady & Rosenthal, 1992, p. 256). Would theories of person perception differ if first impressions and thin slices instead occur at a higher level of awareness than previously assumed (which we as a field would know if we measured awareness in different ways)? Or consider, for another example, research showing that implicit attitudes primarily form and change in response to information presented suboptimally (Rydell et al., 2006). This seminal work argued that converse to explicit attitudes, "implicit attitudes are sensitive to associative information presented below conscious awareness" (p. 957). But how might later attitude theories have otherwise developed if implicit attitude change instead requires a degree of conscious awareness (which we as a field would know if we measured awareness in different ways)? Previous decisions to dichotomize awareness into an all-or-none effect have certainly influenced the direction and development of many psychological theories.

Instead of dichotomizing awareness versus unawareness, researchers may be better served by considering awareness along a continuum from nil to total. Such a gradient view uses the term "(un)awareness" to refer to *measurable levels* along a continuum (Overgaard et al., 2006). As in, participants could be aware that something was presented (i.e., a very low-level awareness), and be able to pick it out of a lineup (i.e., a moderate level of awareness) but be simultaneously unaware of what was presented (i.e., lacking the highest level of awareness). By doing so, the choice of awareness measure is simplified as it depends only on one's goal within a hierarchy of perception. One need only to determine the level of perception one wishes to isolate the influence of and design an awareness measure to match that level. The idea of relative thresholds reflecting unique sensitivities was discussed at length by Snodgrass and colleagues (2004) who highlight a hierarchy of perception based on the amount of information needed to register a correct response. At the bottom of the hierarchy were mere detection tasks (e.g., the ODT task used in the current work), which were proposed to be the most sensitive, followed by identification tasks (e.g., the AFC task used in the current work), then semantic classification or object identification tasks (e.g., the self-report task used in the current work). They argue that "achieving null sensitivity on the relevant lower order task (which requires more stringent exposure conditions) should suffice to prevent conscious, higher order perceptual effects" (p. 853). They rely on earlier work by Macmillan and Creelman (1991) to argue that if conscious perception reflects a hierarchy of perception, awareness measures may therefore reflect a hierarchical relationship. Here, higher level awareness (e.g., self-report) relies on lower level effects (e.g., mere detection). So, without the lower order effects, the higher order effects would not be possible. Time and effort might also be important influences. It is more work to type in a response than simply selecting yes/no. Time needed to respond might result in decay of fragile percepts (Greenwald et al., 1996). The time to respond is a feature that should be matched to one's research design.

These classic psychophysics views on detection versus identification are valuable considerations when conducting consciousness research (Macmillan & Creelman, 2005). Indeed, it has recently been argued that people experience a richer phenomenal consciousness than they can report (Block, 2011). This is why they might do poorly in a self-report task yet show greater sensitivity when asked to make simple decisions like detection or forced-choice identification. This idea is supported by the current findings. Treating all measures of awareness as reflecting the same underlying phenomenon fails to consider how differences among measures reflect important information about the very processes under investigation. From this perspective, any awareness measure can be valid/invalid depending on the level of awareness the researcher wishes to maintain. In other words, the goal of the researcher is paramount to determining the proper awareness measure. Would mere detection of the presentation of a stimulus, absent the ability to identify its affective content or semantic category, alter the measured downstream response to that stimulus? If so, then even this state of awareness may be unacceptable to a researcher's goals. If the goal of suboptimal presentation is to prevent the influence of conscious perception, anything less than a correct self-report can be characterized as unaware and considered acceptable. Alternatively, it may be that different measures actually tap different underlying processes rather than gradients along one construct. Both views present unique issues for measuring awareness, and a comprehensive discussion on the merits of each view is a ripe target of future work. Paramount to the current discussion going forward is a consideration of the goal of the research(er). Consideration of the nature of awareness is a topic for separate work.

In the end, as I stated early on, the appropriateness of an awareness measure is directly tied to a consideration of what purpose limiting awareness is being used for and what level of perception one wishes to allow. That consideration will make clear the *degree or type of awareness* that must remain below the chosen threshold of perception. Through careful consideration of this question, a researcher can identify or design an appropriate awareness measure. The takeaway is clear: awareness as a concept is still debated, yet a strict definition of awareness is generally unhelpful; when one is interested in presenting stimuli "below threshold," it is necessary to consider the specific threshold of perception one wishes to preclude. Once determined, the choice of awareness measure should be simple and easily justified in any resultant research. It should, of course, always be acknowledged that there are limitations in using any particular measure. As a specific example, imagine one wants to largely prevent participants from identifying pictures and to see if those pictures influence choices on a subsequent task. They could implement the methods for a SSR task and would likely be able to show that these pictures resulted in some priming effect or influence on a future choice. They would also need to make sure that readers of their study results were informed that these pictures could probably be detected or partly identified and to couch their results as yet another example of subjective threshold stimulus effects, which are likely conscious to some degree.

CONSERVATIVE DECISION BIAS

Conservative decision bias is a major issue as it can hide true sensitivity to masked stimuli. SDT metrics like c and a'/a are supposed to capture and reflect the combination of sensitivity and bias, yet these metrics can only indicate a bias, not the reason for it, and the reason can be critically important. In the current work on the open response experiments, the "don't know" rates were quite high, at 62.2% (backward mask) and 74.5% (sandwich mask). These rates were not treated herein as a conservative bias as it was instead assumed that the task was very difficult. In the yes/no ODT, c indices indicated a bias towards saying that nothing was presented. Again, this was not considered a "bias" per se, but an accurate reflection of the subjective experience of the participant. That is, they honestly reported that nothing was presented because they did not perceive that anything was presented. This type of bias is particularly troublesome for present/absent tasks and the accuracy-dependent enumeration task used herein as what looks like bias may actually reflect the reality of the participant's experience. Training participants to be sensitive to minimal or "fleeting" cues may help reduce such bias.

Yet people might report "no" when they really did consciously experience something for several reasons, for example, to finish the study sooner or in order to avoid potential embarrassment. Alternatively, the content of a category of stimuli may be a source of fear or anxiety (i.e., spiders to a spider-phobic person), which may encourage psychologically "avoiding" the stimulus. More so, as was seen in the 8-AFC task, not only may people psychologically avoid engaging with certain classes of stimuli, but they may also visually and decisionally avoid them. Here participants appeared much more accurate for neutral and positive than negative and threatening objects. Yet, as quoted earlier, some participants will not want to select violent images or look at the disgusting pictures. This type of response bias leads to inflated accuracy for some categories and deflated accuracy for others. This is an oft-overlooked issue when measuring awareness. Consider the case of Ruiz-Padial and Vila (2007), who administered a recognition task on which participants simultaneously saw samples of several categories of pictures (e.g., spiders, flowers, mutilation) and reported which were presented during the main task. AFC tasks presenting multiple category choices that differ in attractiveness or repulsiveness (or any number of other dimensions) may be unreliable measures of awareness that reflect response bias toward or away from certain classes of stimuli. The consequence of ignoring decision bias may yield results that give a misleading appearance that unconscious processing has or has not been achieved.

ON MEASURING AWARENESS DURING THE MAIN TASK VERSUS A SEPARATE TASK

An area of debate in literature using awareness measures involves whether to measure awareness trial-by-trial during the main task or to measure after the main task. Advocates of online measurement suggest interweaving within the task trials in which the participant indicates awareness. Others have suggested that although an online check may be feasible within certain designs, a trial-by-trial or even intermittent check of what people saw is not possible within many if not most paradigms. This position argues that an online check asking participants to list or match what they saw would interfere with the processes underlying the measured response. That is, bringing people's attention to the suboptimal stimuli can change participants' behavior in the study, and may amplify, undo, or interact with critical effects (Hauser & Schwarz, 2015; Hauser et al., 2018). Recall my previous study employing the startle-eyeblink paradigm where an online rating of what people saw rendered startle eyeblink data completely useless.

In separate work I employed suboptimal stimuli and a valence-rating task where participants rated the goodness-to-badness of each suboptimal stimuli. Asking people to report what they saw on each trial would have interfered with their ability to rate valence based on gut-level affect. The participants' guess of what they saw on every trial would likely contaminate and guide their valence rating, regardless of whether they guessed correctly. That is, participants will begin guessing and engaging in processing they are not *naturally doing*. Meaning, measuring both valence and awareness simultaneously renders the more sensitive valence rating useless. In this case, the awareness measure itself can compromise a physiological measure. The expectation of a separate online awareness check embedded within each measure is not possible as in many cases it would undermine the measure itself. This can be seen as a limitation of certain measures but does not invalidate the conclusion. In such instances, a separate measure of awareness is more appropriate. The researcher can decide a priori how to identify "aware" participants and what exclusion criteria will be used to exclude those participants.

ON IDIOSYNCRATIC ABILITIES

Critical to much suboptimal research is presenting stimuli below the specified threshold of awareness. This would ideally involve at the minimum a pilot of the presentation paradigm to ensure that stimuli are undetectable at the specified threshold. The current work converges with previous work to indicate that individuals vary in their idiosyncratic ability to perceive quickly presented stimuli (Bengson & Hutchinson, 2007; Pessoa et al., 2005). I descriptively explored between-subject differences in ability, and histograms clearly indicate wide-ranging abilities. Participants evidenced a range of individual differences on all measures, as some demonstrated near-chance ability and others were nearly perfectly accurate. Researchers may therefore need to account for inter-individual differences in intrinsic ability to perceive quickly presented stimuli.

Some previous research has attempted to "dial in" stimulus durations on an individual basis by undertaking an awareness calibration prior to the main task, where the stimulus duration is incrementally lowered until an awareness measure indicates that an individual participant is no longer perceiving stimuli above whatever threshold has been chosen (see Tsikandilakis et al., 2023, for a method describing this process [with extensive effort]). Although the fact that participants have a wide range of idiosyncratic skills is certainly a concern, adjusting stimulus durations per participant is often not a viable option for several reasons. In some research designs, participants are never made aware that suboptimal stimuli are being presented, and illuminating this fact during a calibration procedure would spoil the subterfuge. Alternatively, to truly dial in an effective suboptimal duration for a specific set of stimuli, the same stimuli from the main task should be used during the calibration procedure. Yet the calibration process necessarily entails that participants are at some point aware of many of the stimuli. This results in two issues. One is the confound that some participants are more aware of the content of the stimuli than are others. The second is that if concealing the content of the stimuli is important to the main research question, exposing that content during calibration is likely to confound the main study. If these and other issues are not a concern, dialing in stimulus duration on a per-participant basis may be ideal. But this will often not be appropriate for many reasons. In such cases, it may often be more suitable to measure awareness after the main task and exclude people who evince awareness above the desired threshold.

ON THE PREFERENTIAL RESPONDING TO DIFFERENT CLASSES OF EMOTIONAL STIMULI

Although it is not the main focus of the current article, the work that prompted these studies (i.e., March et al., 2022) intended to rule out the possibility that preferential responding to threat stimuli (March et al., 2018a, 2018b, 2020) could be partly explained by between-category differences in masking efficacy. That is, could certain classes of emotional stimuli have superior sensitivity (i.e., be harder to mask, easier to sense) compared to neutral stimuli in masked presentations? It has been shown that low-level visual attributes can be confounded with emotional categories (Lakens et al., 2013). This confound could lead to a response bias that would be important to separate from perceptual sensitivity in order to get an accurate measure of awareness. Indeed, within the current studies, performance varied as a function of the stimulus category. On the ODT, people were equally capable of detecting the presence of threatening versus neutral stimuli, and both were detected better than negative or positive stimuli. Yet in the 3-AFC-SM task, while people were best able to pick the correct neutral alternative, they performed worst when attempting to match threat and negative stimuli, and performance on both was worse than when matching positive stimuli. On the SSR, participants once again performed the best on neutral trials (even if performance was only 12% and 2% accuracy on the BM and SM, respectively), and they generally performed equally poorly on all the other categories. Across all studies, the most consistent trend involved neutral objects as more readily perceived as present, matched, and described than were objects from the other categories and, on the contrary, negative objects as least readily perceived as present, matched, and described.

Three apparent explanations for this effect are between-category differences in the (1) level of masking necessary to achieve an objective threshold for unawareness, (2) perception of activated nonconscious affect, and/or (3) inclination to avoid feeling, reporting, or choosing anything negative. Response sensitivity bias could reflect the influence of one or several of these (and likely other) sources. Differences in performance between categories may reflect something visually particular about the items present in that category. Some classes of stimuli may contain category-canonical features that render them harder to suppress below conscious perception (e.g., sharp edges, higher levels of contrast). Some emotional content may activate nonconscious affect that is perceptible to the participant, and such affect may then be used as information when responding on an awareness measure. Previous work has noted suboptimally presented happy faces are detected better than other emotional faces (Sweeny et al., 2013) and harder to mask (Maxwell & Davidson, 2004). This could seemingly reflect felt affect, or as implied by the 8-AFC task of Study 2 of the current work, participants might be more inclined to respond to a positive emotional response rather than a negative emotional response (or as is the case in the 8-AFC task, to avoid attending to negative alternatives).

Future work employing distinct categories of stimuli may need to account for between-category differences in perceptibility. Furthermore, it has recently been argued that individual stimuli should be treated as random effects in statistical models. That is, it is important know whether even within classes of stimuli, effects are being driven by certain stimuli (Judd et al., 2012), or specific items "explain" perception without awareness (Fisk & Haase, 2007). Future work may explore awareness as a function of the characteristics of specific stimuli to determine whether features that render stimuli harder to present outside awareness are more prevalent within certain classes of stimuli.

CONCLUSION

The current work reports six different data sets, orthogonally combining two types of masking, sandwich masking and backward masking, with three types of tasks, a present/absent yes/no task, a discrimination task, and a present/absent task but with accuracy-dependent enumeration. The results showed biases towards particular judgments, individual variation, less than perfect masking, and maybe task or stimulus dependencies. My goal here was not to suggest that one of these (or other) measures of awareness is *best* or most appropriate. There are several other articles arguing for the use of certain measures. On the contrary, the level of unawareness necessary and the appropriate way to measure awareness depends on one's specific research goals. While I contend that no method is better, I do contend that the appropriate measure should be contingent on the theoretical bases of one's research goals. The argument presented here is that an overall definition of awareness is not possible (given the controversy in the field regarding its measurement); consequently, measuring awareness should depend on the researcher's objective. If a research question requires that participants remain unaware of even the mere presence of a stimulus, a yes/no task is commensurate with such requirements. But on the other hand, such liberal tests of awareness may be incongruent with the typical goal of some suboptimal presentation paradigms: to keep stimuli below the threshold for *conscious* classification. In this situation, any level of awareness below explicit perception may be satisfactory. And there are many levels of awareness between these two poles not tested in the current work that a researcher may be interested in. Considering awareness in studies using suboptimal stimuli requires both knowing what level of awareness the researcher hopes to preclude and the level of awareness evidenced by participants.

REFERENCES

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, *32*, 201–271. https://doi.org/10.1016/S0065-2601(00)80006-4

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256–274. https://doi.org/10.1037/0033-2909.111.2.256

Baier, D., Kempkes, M., Ditye, T., & Ansorge, U. (2022). Do subliminal fearful facial expressions capture attention? *Frontiers in Psychology*, *13*, 840746. https://doi.org/10.3389/fpsyg.2022.840746

Bar-Anan, Y., & Amzaleg-David, E. (2014). The effect of evaluation on co-occurrence memory judgement. *Cognition and Emotion*, *28*(6), 1030–1046. https://doi.org/10.1080/02699931.2013.867835

Bengson, J. J., & Hutchison, K. A. (2007). Variability in response criteria affects estimates of conscious identification and unconscious semantic priming. *Consciousness and Cognition*, *16*(4), 785–796. https:/doi.org/10.1016/j.concog.2006.12.002

Bevan, W. (1964). Subliminal stimulation: A pervasive problem for psychology. *Psychological Bulletin*, *61*, 81–99. https://doi.org/10.1037/h0048670

Bitbol, M., & Petitmengin, C. (2017). Neurophenomenology and the micro-phenomenological interview. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 726–739). John Wiley & Sons.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition*, *35*, 234–250. https://doi.org/10.1016/j.concog.2015.01.012

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision.* Oxford University Press.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*(4), 387–395. https://doi.org/10.3758/bf03202793

DeCarlo, L. T. (2012). On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, *56*(3), 196–207. https://doi.org/https://doi.org/10.1016/j.jmp.2012.02.004

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Di Lollo, V. (2018). Attention is a sterile concept; iterative reentry is a fertile substitute. *Consciousness and Cognition*, *64*, 45–49. https:/doi.org/10.1016/j.concog.2018.02.005

Dienes, Z., & Perner, J. (2004). Assumptions of a subjective measure of consciousness. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology* (pp. 173–199). John Benjamins.

Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition*, *19*(2), 674–681. https://doi.org/10.1016/j.concog.2009.09.009

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, *67*, 279–300. https://doi.org/10.1037/h0041622

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115–141. https://doi.org/10.1080/0269993004200024

Fisk, G. D., & Haase, S. J. (2005). Unconscious perception or not? An evaluation of detection and discrimination as indicators of awareness. *American Journal of Psychology*, *118*(2), 183–212.

Fisk, G. D., & Haase, S. J. (2007). Exclusion failure does not demonstrate unconscious perception. *American Journal of Psychology, 120*(2), 173–204.

Gelbard-Sagiv, H., Faivre, N., Mudrik, L., & Koch, C. (2016). Low-level awareness accompanies "unconscious" high-level processing during continuous flash suppression. *Journal of Vision*, *16*(1), 3. https://doi.org/10.1167/16.1.3

Goldiamond, I. (1958). Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychological Bulletin*, *55*(6), 373–411. https://doi.org/10.1037/h0046992

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273*(5282), 1699–1702. https://doi.org/10.1126/science.273.5282.1699

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology, 9*, 998. https://doi.org/10.3389/fpsyg.2018.00998

Hauser, D. J., & Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *Sage Open, 5*(2), 2158244015584617. https://doi.org/10.1177/2158244015584617

Hedger, N., Gray, K. L. H., 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. *Psychological Bulletin, 142*(9), 934–968. https://doi.org/10.1037/bul0000054

Irvine, E. (2012). Old problems with new measures in the science of consciousness. *British Journal for the Philosophy of Science*, *63*(3), 627–648.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science*, *237*(4821), 1445–1452.

Koch, C., & Preuschoff, K. (2007). Betting the house on consciousness. *Nature Neuroscience, 10*(2), 140–141. https:/doi.org/10.1038/nn0207-140

Koster, N., Mattler, U., & Albrecht, T. (2020). Visual experience forms a multidimensional pattern that is not reducible to a single measure: Evidence from metacontrast masking. *Journal of Vision*, *20*(3), 2. https:/doi.org/10.1167/jovi.20.3.2

Lähteenmäki, M., Hyönä, J., Koivisto, M., & Nummenmaa, L. (2015). Affective processing requires awareness. *Journal of Experimental Psychology: General*, *144*(2), 339–365. https:/doi.org/10.1037/xge0000040

Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion, 27*, 1225–1246. https://doi.org/10.1080/02699931.2013.781501

Lohse, M., & Overgaard, M. (2019). Emotional priming depends on the degree of conscious experience. *Neuropsychologia, 128,* 96–102. https://doi.org/10.1016/j.neuropsychologia.2017.10.028

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Mangan, B. (2001). Sensation's ghost: The non-sensory "fringe" of consciousness. *Psyche*, *7*(18).

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*(2), 197–237. https://doi.org/10.1016/0010-0285(83)90009-9

March, D. S., Gaertner, L., & Olson, M. A. (2017). In harm's way: On preferential response to threatening stimuli. *Personality and Social Psychology Bulletin*, *43*(11), 1519–1529. https://doi.org/10.1177/0146167217722558

March, D. S., Gaertner, L., & Olson, M. A. (2018a). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry, 29*(1), 1–13. https://doi.org/10.1080/1047840X.2018.1435680

March, D. S., Gaertner, L., & Olson, M. A. (2018b). Clarifying the explanatory scope of the Dual Implicit Process model. *Psychological Inquiry, 29*(1), 38–43. https://doi.org/10.1080/1047840X.2018.1435622

March, D., S., Olson, M., A., & Gaertner, L. (2020). Lions, and tigers, and implicit measures, oh my! Implicit assessment and the valence vs. threat distinction. *Social Cognition, 38*(Suppl), S154–S164. https://doi.org/10.1521/soco.2020.38.supp.s154

March, D. S., Gaertner, L., & Olson, M. A. (2022). On the automatic nature of threat: Physiological and evaluative reactions to survival-threats outside conscious perception. *Affective Science*, *3*(1), 135–144. https://doi.org/10.1007/s42761-021-00090-6

Maxwell, J., & Davidson, R. (2004). Unequally masked: Indexing differences in the perceptual salience of "unseen" facial expressions. *Cognition & Emotion*, *18*(8), 1009–1026.https://doi.org/10.1080/02699930441000003

Merikle, P. M. (1984). Toward a definition of awareness. *Bulletin of the Psychonomic Society*, *22*, 449–450. https://doi.org/10.3758/BF03333874

Merikle, P. M., & Cheesman, J. (1987). Current status of research on subliminal perception. *Advances in Consumer Research, 14*, 298–302.

Michel, M. (2023). How (not) to underestimate unconscious perception. *Mind & Language, 38*(2), 413–430. https://doi.org/10.1111/mila.12406

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*(1), 1–19. https://doi.org/10.1017/S0140525X12003214

Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Consciousness and Cognition*, *15*(4), 700–708. https://doi.org/10.1016/j.concog.2006.04.002

Overgaard, M., & Sandberg, K. (2021). The Perceptual Awareness Scale—Recent controversies and debates. *Neuroscience of Consciousness, 2021*, niab044. https://doi.org/10.1093/nc/niab044

Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion*, *5*(2), 243–247. https://doi.org/10.1037/1528-3542.5.2.243

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*(1), 1–23. https://doi.org/10.1023/B:PHEN.0000041900.30172.e8

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language*, *5*(1), 9–28. https://doi.org/10.1111/j.1468-0017.1990.tb00150.x

Rothkirch, M., & Hesselmann, G. (2017). What we talk about when we talk about unconscious processing—A plea for best practices. *Frontiers in Psychology, 8*, 835. https://doi.org/10.3389/fpsyg.2017.00835.

Ruiz-Padial, E., & Vila, J. (2007). Fearful and sexual pictures not consciously seen modulate the startle reflex in human beings. *Biological Psychiatry*, *61*(8), 996–1001. https://doi.org/10.1016/j.biopsych.2006.08.046

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science, 17*, 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, *19*(4), 1069–1078. https://doi.org/10.1016/j.concog.2009.12.013

Schmidt, T., & Vorberg, D. (2006). Criteria for unconscious cognition: Three types of dissociation. *Perception & Psychophysics, 68*(3), 489–504. https://doi.org/10.3758/bf03193692

Sidis, B. (1898). The psychology of suggestion. *Science*, *8*(188), 162–163.

Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception: A model-based approach to method and evidence. *Perception & Psychophysics, 66*(5), 846–867. https://doi.org/10.3758/bf03194978

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/bf03207704

Stein, T., & Peelen, M. V. (2021). Dissociating conscious and unconscious influences on visual detection effects. *Nature Human Behaviour, 5*(5), 612–624. https://doi.org/10.1038/s41562-020-01004-5

Storbeck, J. & Clore, G. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass, 2*(5), 1824–1843. https://doi.org/10.1111/j.1751-9004.2008.00138.x

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion*, *13*(1), 76–91. https://doi.org/10.1037/a0029193

Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of the current methods. In M. Overgaard (Ed.), *Behavioral methods in consciousness research* (pp. 21–46). Oxford University Press.

Tsikandilakis, M., Bali, P., Karlis, A., Mével, P. A., Madan, C., Derrfuss, J., & Milbank, A. (2023). Unbiased individual unconsciousness: Rationale, replication and developing applications. *Current Research in Behavioral Sciences*, *4*, 100109. https://doi.org/10.1016/j.crbeha.2023.100109

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,* 432–450. https://doi.org/10.1177/0301006618760738

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences, 19*(12), 757–770. https://doi.org/10.1016/j.tics.2015.10.002

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. *Biological Psychology, 129,* 131–153. https://doi.org/10.1016/j.biopsycho.2017.08.051

Wiens, S. (2005). Remain aware of awareness. *Nature Reviews Neuroscience, 6*(3), 258. https://doi.org/10.1038/nrn1630-c1

Wiens, S. (2006). Current concerns in visual masking. *Emotion, 6*(4), 675–680. https://doi.org/10.1037/1528-3542.6.4.675

Wiens, S., & Öhman, A. (2002). Unawareness is more than a chance event: Comment on Lovibond and Shanks (2002). *Journal of Experimental Psychology: Animal Behavior Processes, 28*(1), 27–31. https://doi.org/10.1037/0097-7403.28.1.27

Wierzchoń, M., Asanowicz, D., Paulewicz, B., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and Cognition, 21*(3), 1141–1153. https://doi.org/10.1016/j.concog.2012.05.012

Wierzchoń, M., Paulewicz, B., Asanowicz, D., Timmermans, B., & Cleeremans, A. (2014). Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings. *Consciousness and Cognition, 27,* 109–120. https://doi.org/10.1016/j.concog.2014.04.009

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

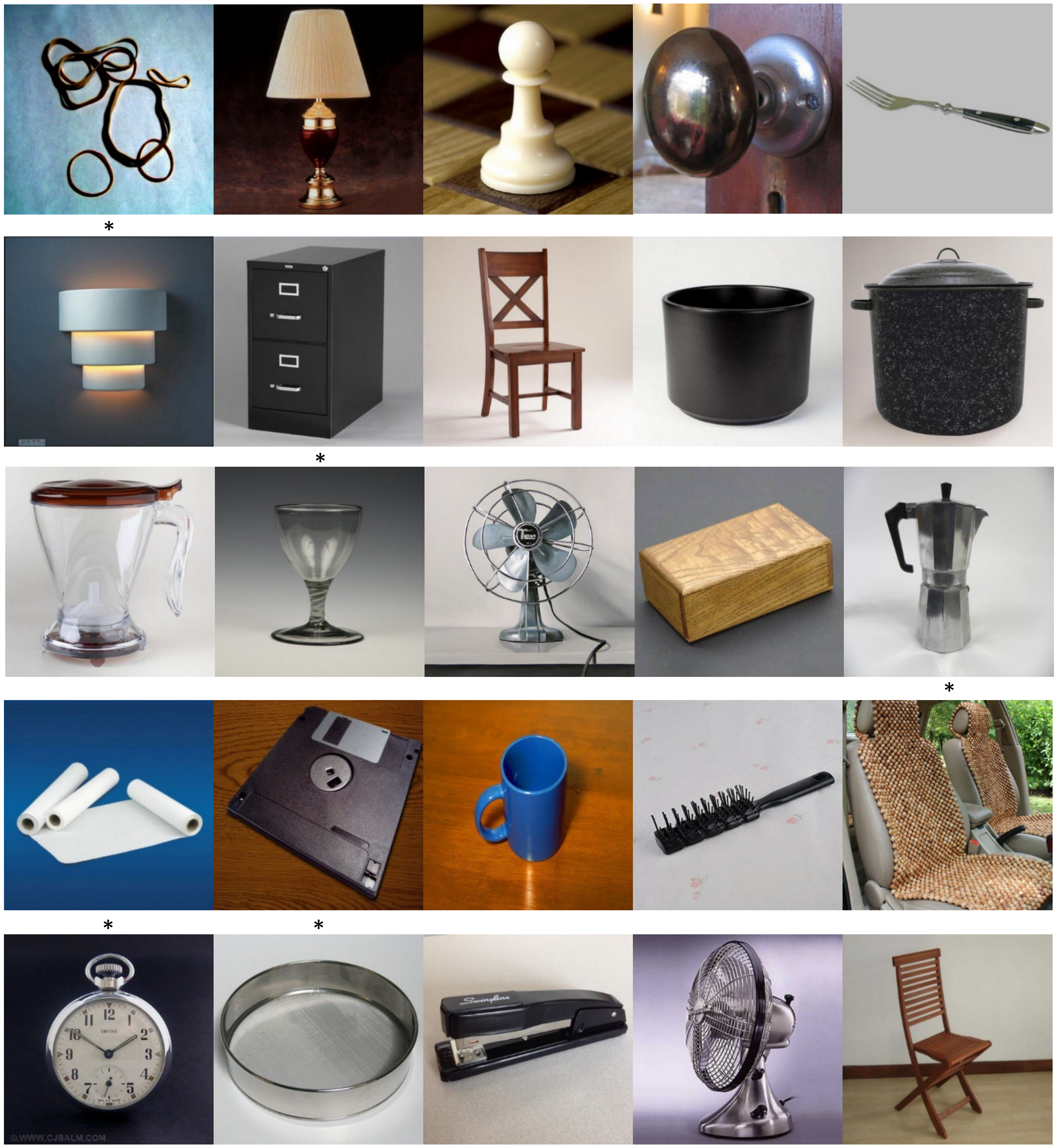
Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 1–10. https://doi.org/10.1007/s11336-003-1119-8

SUPPLEMENTAL MATERIAL

Threat Stimuli

Negative Stimuli

Neutral Stimuli



Positive Stimuli