THEORETICAL NOTE

Current Concerns in Visual Masking

Stefan Wiens
Stockholm University and Karolinska Institute

Theories of emotion postulate that emotional input is processed independently from perceptual awareness. Although visual masking has a long tradition in studying whether emotional pictures are processed below a supposed threshold of perceptual awareness (subliminal perception), a consensus has yet to be reached. This article reviews current concerns in the use of visual masking. These include a reliable presentation method, the role of masking pictures, common definitions of awareness and their problems, current models of awareness, and neural mechanisms. A useful strategy may be the study of dose-response relationships between awareness and emotion processing that avoids a dichotomous view of awareness and allows conclusions about the relative independence of emotional processing from awareness.

Humans laugh at jokes, cry in desperation, and run from the proverbial bear. Because many of these responses are elicited spontaneously and without any apparent effort, emotions appear to occur automatically. Indeed, because consciousness has limited capacity (Morais & Ivanoff, 2005; Roser & Gazzaniga, 2004; Shevrin & Dickman, 1980), theories of emotion suggest that emotional stimuli are processed independently from perceptual awareness, particularly so in detecting and responding to potential threat (Dolan & Vuilleumier, 2003; LeDoux, 2000; Öhman, 1986; Robinson, 1998).

Perceptual awareness needs to be manipulated to study its role in processing emotional pictures. Visual masking is a widely used method to manipulate perceptual awareness (Holender, 1986; Kim & Blake, 2005). For example, in backward masking, an emotional picture (target) is shown briefly and followed quickly by another picture (mask). If picture parameters are adjusted carefully, people report that they are aware only of the mask but not the target. For example, when fearful and happy faces (targets) are masked by neutral faces (masks), people might report that they are not aware of the emotional facial expressions. The question of whether emotional facial expressions are processed independently from perceptual awareness and attention has been studied extensively (Morris, Öhman, & Dolan, 1998; Pessoa, Japee, Sturman, & Ungerleider, 2006; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Philips et al., 2004; Vuilleumier, Armony, Driver, & Dolan, 2001; Whalen et al., 1998). However, no consensus has been reached (Pessoa, 2005; Vuilleumier, 2005). The purpose of this article is to review current (and recurrent) concerns in the use of visual masking to study the role of perceptual awareness in emotional picture processing.

Reliable Presentation Method

An unreliable masking procedure introduces unwanted variability in stimulus parameters over trial repetitions. Because research on visual masking suggests that small changes in stimulus parameters (e.g., duration) result in large changes in perception (e.g., Esteves & Öhman, 1993), an unreliable procedure causes difficulty in maintaining particular levels of perception and in ensuring that changes in perception result from internal perceptual processes rather than from external variations in stimulus parameters. Unfortunately, my colleagues and I found that different display technologies generally tend to be more unreliable than claimed by manufacturers and assumed by researchers (Wiens et al., 2004). On the basis of these results, researchers are advised to provide evidence for the reliability of their displays (for a masking setup in imaging, see Wiens & Öhman, 2005b). Also, because different display technologies have different effects on stimulus parameters (e.g., luminance curves), comparing specific parameters (e.g., duration) between studies that used different displays may be of limited value. Finally, because perception varies with small changes in stimulus parameters, it is important to keep testing conditions as consistent as possible over time and participants. However, variables such as background illumination, viewing angle, and dark adaption might be difficult to equate completely, particularly so in brain imaging research. Because these variables introduce undesired “noise” in the data and may confound effects from other variables of interest (e.g., state anxiety), these considerations suggest that perceptual awareness ought to be assessed rather than assumed for each participant. If concerns exist that the measurement of awareness itself might bias results (cf. Whalen et al., 1998), awareness might be assessed in a separate task.

Role of Masking Pictures

In many studies, masks are seen only as means to manipulate perceptual awareness of the targets, and potential effects from their
content are neglected (Eriksen, 1980). For example, it is common practice to use neutral facial expressions to mask emotional facial expressions (Morris, Öhman, & Dolan, 1998; Pessoa, Japee, Sturm, & Ungerleider, 2006; Phillips et al., 2004; Whalen et al., 1998, 2004). Because emotional expressions can be masked more easily with pictures of faces than with other pictures (Costen, Shepherd, Ellis, & Craw, 1994), faces may have been used mainly out of necessity to overcome technical limitations in presenting emotional target faces sufficiently briefly. However, a consequence of this practice is that results cannot be unequivocally attributed to effects from the targets per se. Instead, results may occur for only particular combinations of targets and masks. For example, because the amygdala responds more strongly to motion (LaBar, Crupain, Voyvodic, & McCarthy, 2003), it is possible that amygdalar activation to masked faces is not a direct effect of the emotional targets but is due to a motion effect from transitions from emotional target faces to neutral mask faces. Although it has been suggested that motion effects may be eliminated by offsetting the position of target and mask (Phillips et al., 2004), ruling out effects from masks appears difficult unless the masks themselves do not elicit any responses. However, although using masks that lack a central object (e.g., scrambled faces) may be possible, this procedure may change perceptual processes. Whereas participants might not suspect targets when these are masked by meaningful pictures (e.g., neutral faces), they might search actively for masked targets with masks that lack a central object (e.g., scrambled faces). Research is needed to study effects from different masks in visual masking.

For masking to be valid, it is further necessary that masking work equally well for all target categories and for all target pictures within a category (Eriksen, 1980). If pictures from different target categories differ in their visibility, category effects might be due to confounding effects from perception. Whereas these requirements are often assumed, research suggests that different target categories may not be masked alike. For example, neutral faces tend to mask happy faces less well than angry faces (Maxwell & Davidson, 2004). This finding suggests that researchers need to study whether perceptual awareness differs for different target–mask categories. However, it is important to consider that any differences may not necessarily reflect perceptual confounds but genuine findings. Thus, target categories might have different perceptual consequences due to their emotional character (Reinders, den Boer, & Buchel, 2005; Vuilleumier & Schwartz, 2001). Heterogeneity of masking within a target category may also confound results. For example, if masking is effective for all but a few targets, any emotional effects might be due to these targets, or participants might discriminate targets better than chance only because they can discriminate these few targets while purely guessing on the remaining masked targets. Heterogeneity of masking may be assessed by computing how many times particular targets are detected across trials and participants. If the number of detected pictures is comparable among targets and shows a narrow range, these findings would be evidence for homogeneity of picture visibility among targets.

Unawareness as Chance-Level Discrimination Ability

In most research with visual masking, the goal has been to determine whether emotional pictures are processed even if participants are not perceptually aware of the pictures (Merkle, Smilek, & Eastwood, 2001). To study whether perception occurs below the limen or threshold of perceptual awareness (subliminal perception), elimination of perceptual awareness is necessary. If responses occur after eliminating perceptual awareness, awareness cannot be a necessary condition for their occurrence (Frith, Perry, & Lumer, 1999). Because few, if any, challenge the assumption that participants are unaware if they cannot discriminate between emotional and unemotional pictures, absence of perceptual awareness often has been indexed as the level at which participants cannot discriminate between the pictures. With this focus on behavioral performance regardless of self-reported level of awareness (Eriksen, 1960), this objective threshold is distinguished commonly from the subjective threshold that indexes participants’ self-reported level of awareness (Cheesman & Merkle, 1984).

In measuring the objective threshold, many researchers advocate the use of signal-detection measures. A critical feature of signal-detection theory is that discrimination ability is separated from response bias, that is, participants’ tendency to favor a particular response alternative (Macmillan & Creelman, 1991). For example, to determine whether participants can discriminate fearful from neutral faces, researchers may show fearful and neutral faces and ask participants to respond on every trial whether they detected a fearful face (yes or no). Proportions of yes responses on trials with a fearful face (hits) can be compared with proportions of yes responses on trials with a neutral face (false alarms). The greater the proportion of hits than false alarms, the greater the ability to discriminate between fearful and neutral faces. Critically, this discrimination ability is independent from participants’ response biases. That is, both hits and false alarms are affected by how much participants favor to respond yes (that they detected a fearful face). In signal-detection theory, discrimination ability may be indexed by $d'$ and response bias by $C$ (Macmillan & Creelman, 1990; Snodgrass & Corwin, 1988). From this perspective, the objective threshold corresponds to the energy threshold or an empirical threshold that is set at $d' = 0$ (Macmillan, 1986). Although this is the most conservative approach to indexing perceptual unawareness, there are several problems with it.

Because the claim for an objective threshold ($d' = 0$) is an attempt to prove the null hypothesis, it requires strong statistical power to reduce the risk for a Type II error (falsely concluding that $d' = 0$). Power increases with number of observations and decreases with variability among observations (Rosenthal & Rosnow, 1991). Accordingly, if the number of observations (trials or participants) is small, a considerable effect size might not result in significance. This is illustrated by the counternull value of an effect (Rosenthal & Rubin, 1994). The counternull value is the upper limit of performance that is equally supported by the data as the null hypothesis. Thus, both null and counternull are associated with the same significance levels. For example, if a study reports that mean performance of $d' = 0.8$ did not differ significantly from zero ($p = .15$), then the counternull value is $2 \times 0.8 = 1.6$. This result means that although the observed mean performance of $d' = 0.8$ does not differ significantly from a population $d' = 0$ ($p = .15$), it does not differ significantly from a population $d' = 1.6$, either ($p = .15$). Because $d' = 1.6$ has considerable size, this example illustrates that nonsignificance does not necessarily indicate zero effect. Similarly, because power decreases with increased variability among observations, variability among participants...
may result in nonsignificance. Indeed, findings of large individual differences would suggest that mean performance may not be a representative index for the group. To evaluate individual differences and their effects, confidence intervals as well as correlations with variables of interest (e.g., Whalen et al., 2004) can be calculated (Cumming & Finch, 2005).

Therefore, if researchers intend to argue that \( d' = 0 \), they need to provide convincing evidence for that. Although statistical power can be obtained with hundreds of trials and a lax significance criterion (e.g., \( \alpha = .20 \)), it is doubtful that participants will stay motivated during this task because signal and no-signal trials will need to be indistinguishable to them (Merikle, 1992). As a result, \( d' = 0 \) may accurately reflect not absence of discrimination ability but lack of motivation (Merikle & Daneman, 2000). Also, the task needs to be sensitive to the relevant stimulus dimension (Duncan, 1985). For example, the task used by Morris et al. (1998) might be criticized because it assessed whether participants could detect target faces rather than whether they could discriminate between them (which was the relevant stimulus dimension). However, even if the relevant stimulus dimension is assessed, which specific task should be used may be unclear. For example, if participants are required to respond whether they saw a fearful or neutral face, this task may be too insensitive if participants can discriminate between the expressions without being able to explicitly label them (cf. Lovibond & Shanks, 2002). Because in brain imaging the evidence for \( d' = 0 \) often has been claimed rather than tested, the conclusion that imaging studies of visual masking are based on criterion (e.g., Whalen et al., 2004) can be calculated from self-reports after the experiment or as performance estimates across many trials. These tasks have limited validity, in that participants are required to keep track of their awareness in memory or have to estimate their awareness by integrating many trials.

Minimum requirements for a valid measure ought to be that it is based on each trial and that the delay between picture presentation and assessment is minimal (Lovibond & Shanks, 2002). A possible candidate for a subjective measure is to instruct participants to report on every trial whether or not they were aware of the picture. In signal-detection theory, this corresponds to the criterion placement and is thus an index of response bias that is neutral about perceptual awareness (Macmillan, 1986). Also, because participants might be instructed to report their level of awareness on a 5-point scale rather than as a yes or no decision, it appears arbitrary to decide about the level below which participants might be considered to be unaware of the targets. Furthermore, even if participants might be instructed to place their criterion at a particular subjective threshold (e.g., participants respond yes only when they are aware of central features of a fearful face), a drawback is that it is unclear how this measure can be averaged over trials. For example, if participants respond yes (i.e., they detected a fearful face) on 20% of the fearful-face trials (hit rate), it is unclear whether they should be considered aware or unaware. Unfortunately, the false-alarm rate (e.g., participants respond yes, they detected a fearful face, on neutral-face trials) cannot be used to make inferences about participants’ subjective awareness, because doing so would assess merely their discrimination ability (as \( d' \) is calculated from hits and false alarms).

An approach with a long tradition in experimental psychology (Merikle et al., 2001) is to consider only signal trials of which...
participants report not to be aware (misses; e.g., no responses on fearful-face trials). However, signal-detection theory can account for findings that missed signals can be discriminated (Haase, Theios, & Jenison, 1999; Macmillan, 1986). The reason is that without considering false alarms, misses reflect only the (arbitrary) placement of the criterion and thus response biases. Accordingly, participants with excellent discrimination ability but with conservative response biases (i.e., they tend to respond no rather than yes) would have a large number of misses, but their discrimination ability would indicate that signals are processed considerably, even for misses. Thus, signal-detection theory permits that when participants are shown masked fearful and neutral faces, they might not report being aware of any faces (misses). Nonetheless, because a miss allows for pictures to be processed considerably, participants might be able to discriminate faces in their expression (fearful compared with neutral), even on misses.

Qualitative Differences

A valid measure of awareness needs to fulfill criteria of exhaustiveness and exclusiveness—meaning that it captures all aspects of perceptual awareness (exhaustiveness) and only those aspects (exclusiveness) (Merikle & Reingold, 1998; Reingold & Merikle, 1990). To illustrate, if a measure is not sensitive enough to capture awareness completely, any emotional effects could be due to the residual aspects of awareness that were not captured by the (inexhaustive) measure. Similarly, if a measure is too sensitive and captures additional aspects outside of perceptual awareness, absence of emotional effects in the absence of perceptual awareness could be due to this (nonexclusive) measure. Unfortunately, there is no agreement on which, if any, measure fulfills these requirements.

Merikle and his colleagues proposed that this debate might be avoided by findings of qualitative differences (for review, see Merikle & Daneman, 2000). Because effects in the absence of perceptual awareness would generally be expected to be weaker than effects from perceptual awareness, an indication for qualitative differences may be the degree to which differences between purportedly aware (conscious) and unaware (unconscious) effects differ from this expected pattern. For example, if results suggest that unconscious effects are stronger than conscious effects (Rotteveel, de Groot, Geutskens, & Phaf, 2001), or are in the opposite direction (Merikle & Joordens, 1997), then these findings would support the distinction between these effects and the validity of the index of awareness. These different effects might be observed at an arbitrary level of performance (e.g., $d' = 1.5$). However, no matter how surprising the findings of apparent qualitative differences, these findings show only that the awareness measure is a correlate of these differences and do not demonstrate a causal relationship (Kunimoto, Miller, & Pashler, 2001).

Prospects

Research on visual masking and awareness needs to consider their neural mechanisms. For example, research suggests that unconscious processes may not necessarily have to be limited to only subcortical mechanisms. Studies in patients with blindsight and neglect and studies of visual masking suggest that substantial cortical activation occurs even though perceptual awareness might be severely degraded or even absent (Goebel, Muckli, Zanella, Singer, & Stoerig, 2001; Haynes, Driver, & Rees, 2005; Vuilleumier et al., 2002). Thus, cortical activation is not sufficient for perceptual awareness. Indeed, recent models of perceptual awareness postulate that it is not cortical processing per se but a certain type of cortical processing (recurrent processing) that is crucial for perceptual awareness (Bullier, 2001b; Enns & Di Lollo, 2000; Lamme, 2003). Thus, masking may not eliminate cortical processing per se but affect only recurrent processing (Bullier, 2001a; Pascual-Leone & Walsh, 2001). An additional point of recent models is that they distinguish between two levels of perceptual awareness, phenomenology (reflective) and awareness (reflective) (Block, 2005; Lamme, 2004). These models postulate that awareness is a higher order process that is secondary to phenomenology but is needed for self-reported awareness (for application to emotional experience, see Lambie & Marcel, 2002; Wiens, 2005). Because awareness (and thus self-report) is neither necessary nor sufficient for phenomenology, phenomenology may occur in the absence of awareness. To illustrate, people might experience mind wandering (phenomenology) without being aware of it (Schooler & Schreiber, 2004). These models sensitize the researcher to the issue that self-report and button presses are only indirect measures of phenomenology. However, if phenomenology is monitored immediately after each trial (as recommended here), dissociations between phenomenology and self-reported awareness are probably minimal. Also, it seems difficult to validate these models externally if self-report is eliminated as a criterion, as there would be no external criterion other than recurrent processing itself.

The issue of awareness in terms of objective discrimination ability or subjective state may be resolved only with a compromise: To acknowledge that both refer to important aspects of perceptual processing, that research needs to study both, and this research also needs to address how concurrent awareness measures affect results. A useful strategy may be to study dose–response relationships between awareness and emotion processing (Wiens & Öhman, 2005a). Dose–response relationships emphasize the continuous nature of awareness and thus encourage researchers to consider effects at various levels of awareness and to characterize awareness with various indexes (e.g., discrimination ability, self-report). Because research on dose–response relationships avoids the current debate in dichotomizing awareness and identifying (arbitrary) thresholds, it may further researchers’ understanding of the role of awareness in emotion.

References


Cowey, A. (2004). The 30th Sir Frederick Bartlett lecture: Fact, artefact,


