Stefan Wiens
Arne Öhman

Probing Unconscious Emotional Processes
On Becoming a Successful Masketeer

Have you ever had an unconscious emotion? As conscious beings, humans might consider this question absurd. Because everyday experience suggests that all important mental events are accompanied by consciousness, humans are convinced that consciousness is critical in mediating and controlling these events and behavior. Accordingly, emotion is commonly equated with emotional experience (feeling), and the notion of unconscious emotion seems either contradictory in terms or not worth discussing, as unconscious emotions would play only a trivial, if any, role at all in mental life. William James (1884), one of the forefathers of modern psychology, supported this notion by stating that “our feeling ... IS the emotion” (p. 190). This discussion of emotion in terms of conscious experience (feeling) has a long tradition in psychology (e.g., Damasio, 1994; Frijda, 1986; Schachter & Singer, 1962). Thus Clore’s (1994, p. 290) statement that “emotions cannot be unconscious because they must be felt, and feelings are by definition conscious” is endorsed by many contemporary emotion researchers.

However, other researchers have challenged this notion of equating emotions with feelings. In fact, this alternative view has a tradition that is at least as long. For example, according to Sigmund Freud (1916–1917/1959), a contemporary of William James’s, unconscious emotion plays an important role in mental life. However, although Freud’s ideas were developed further in psychoanalysis, other researchers have conceptualized unconscious emotions in ways that have little in common with Freud’s theorizing (e.g., Kihlstrom, 1999; LeDoux, 1996; Öhman, 1986).

The purpose of this chapter is to review conceptual, methodological, and technical issues in studying unconscious emotion. We begin with a historical overview of research on unconscious emotion and outline an information-processing model of emotion that illustrates different approaches to studying unconscious emotion. That is, unconscious emotion occurs either when people are not consciously aware of the emotional stimulus or when they show signs of emotion (e.g., psychophysiological changes) even though they do not report any accompanying changes in emotional experience. Then we describe backward masking, which is considered to be the most prominent method for blocking conscious awareness of visual stimuli. In backward masking, a brief visual stimulus (target) is followed immediately by another visual stimulus (mask). Because people often report that they are consciously aware only of the masks but not the targets, masking allows one to reduce, if not eliminate, conscious awareness of masked targets to study their unconscious effects. Based on an example study (Öhman & Soares, 1994), we describe the method and general rationale for using backward masking. Then we discuss technical issues of masking, followed by methodological and conceptual issues in the assessment of awareness. These include the distinction between subjective and objective thresholds, the definition of unawareness in terms of a statistical criterion, and the timing of measuring awareness. Finally, we describe two paradigms that might represent potential solutions to many conceptual problems. Whereas findings of qualitative differences support

the distinction between conscious and unconscious emotional processes, a psychophysical approach offers a more eclectic perspective, as dose-response relationships with awareness can be studied and the use of complementary measures of awareness is advocated. Thus the demonstration of qualitative differences and the study of dose-response relationships between awareness and emotional measures of interest might provide useful tools in probing unconscious emotional processes.

**A Historical Sketch of Unconscious Emotion**

**Sigmund Freud**

In contrast to our everyday experience and traditional conceptualizations of emotion, Sigmund Freud (1916–1917/1959) argued that unconscious processes play an important role in emotion, in that people can be influenced by unconscious emotional and motivational states. For example, emotions might be isolated from their real sources through displacement or projection, as when one's own anger is attributed to one's partner. Similarly, people might be influenced by emotional processes even though defense mechanisms such as denial, intellectualization, and reaction formation might prevent people from experiencing feelings. Thus emotions have to be inferred from their effects on behavior rather than from self-reports of emotional experience. Because feelings can be distorted, if not eliminated, conscious experience is not a necessary component of emotion.

Whereas culture and art of the twentieth century embraced these ideas, academic psychology was skeptical. When psychology became an independent discipline in the nineteenth century, introspection was used to study consciousness, which meant that unconscious processes were neglected. Because behaviorism replaced consciousness with behavior as the object of study, concepts of unconscious, as well as conscious, processes became irrelevant. Further, attempts to introduce psychodynamically inspired notions into mainstream psychology—such as the New Look in perception (Bruner & Goodman, 1947)—were subjected to rigorous criticism (e.g., Erikson, 1960; Holender, 1986; see Ohman, 1999, for review).

**The Cognitive Revolution**

The cognitive revolution of the 1960s and 1970s reintroduced mental phenomena as a primary object for psychologists to study. However, researchers in cognition accepted the behaviorist dictum that mental phenomena had to be studied in behavioral terms (e.g., Mandler, 1975). Accordingly, rather than asking their participants directly to report on mental experience, the cognitive movement sought to reconstruct the architecture of the mind from observing behavioral responses in experimental settings. Thus introspective experience and insights were irrelevant to cognitive theory.

However, with the maturing of cognitive science, the problem of consciousness could be addressed with experimental data (e.g., Posner & Snyder, 1975; Tulving & Schacter, 1990). In this research, consciousness was presumed to have limited resources in that it is slow and can process only one thing at a time (e.g., Posner & Snyder, 1975). However, in addition to the capacity-limited processing of consciousness, another level of information processing was necessary to explain how organisms can process the vast input of external and internal stimuli. For the system to function efficiently, this other level had to process information in an automatic, parallel mode (e.g., Schneider, Dumais, & Shiffrin, 1984). Without this automatic level of information processing, no one would be able to “walk and chew gum at the same time” (to paraphrase a U.S. president’s remark about a colleague). In fact, this perspective challenges the commonsense idea that consciousness plays a critical role in initiating and controlling action. For example, Gazzaniga (1998) argued that the brain works mostly automatically and leaves it to consciousness to interpret and make sense of what is going on after actions have already been executed (see also Roser & Gazzaniga, 2004).

Because information-processing models assign a limited role to consciousness, Shervin and Dickman (1980) argued that the existence of unconscious mental activity seems to be necessary in all psychological models. Indeed, cognitive theories implied unconscious processing stages (Erdelyi, 1974), and, eventually, notions of a cognitive unconscious were introduced (Kihlstrom, 1987). Although most research focused on cognition, researchers began to study unconscious processes in emotion. For example, Zajonc (1980, p. 151) argued that “preferences need no inferences,” suggesting that immediate and automatic emotional reactions to stimuli can occur with a minimum of cognitive processing and can shape subsequent cognitive activity. Similarly, Ohman (1986) proposed that emotional responses to evolutionarily fear-relevant stimuli could occur unconsciously. In most of this research, the method of visual masking was used to probe unconscious emotional processes.

**Delineating Unconscious Emotion**

**Beyond Emotional Experience**

Whereas psychologists interested in cognition accepted that their primary goal was not to elucidate conscious experience but to develop a theory of mental processes that would account for a wide variety of data, including, eventually, conscious experience, most emotion researchers have focused on the experience of emotion (e.g., Ellsworth & Scherer, 2003; Frijda, 1986). However, LeDoux (1996) argued that the science of emotion would profit from following research
on cognition. Accordingly, emotion research should opt for theories that account for systematic bodies of data rather than giving primacy to accounting for subjective experience of emotion. Indeed, a definition of emotion that includes but is not limited to conscious experience is necessary for linking emotion research to other domains of science. For example, an evolutionary perspective on emotion (Ohman & Wiens, 2003; Tooby & Cosmides, 1990) provides a functional scenario in which commonalities and differences between humans and nonhumans can be used to further a general understanding of emotion and to elucidate brain-behavior relationships in emotion (e.g., LeDoux, 1996).

**Inferring Emotion From Four Indicators**

Peter Lang (1968) argued that emotion can be inferred from three different response systems: verbal reports, psychophysiological responses, and behavior. However, these response systems should not be viewed as alternative indicators of a unitary emotional state (presumably mirroring experience) but rather as a loosely coupled ensemble of partially correlated outputs that may show discordant changes to experimental manipulations (Lang, 1993). Indeed, because each response system reflects many sources of variance in addition to emotionally relevant ones, there are limits to their possible covariation. As a result, dissociation between them is a mathematical necessity (Ohman, 1987).

Although an emotion incorporates changes in verbal, psychophysiological, and behavioral components, it also requires an eliciting stimulus, the evaluation of which (as good or bad) is the essential function of the emotion (e.g., Oatley & Jenkins, 1996). Accordingly, we are angry about something, afraid of somebody, and happy because of some event. Theorists of widely different orientations agree that the perceived (or attributed) cause of the emotion is a central determinant of emotional experience (e.g., Damasio, 1994; Schachter & Singer, 1962; Smith & Ellsworth, 1985). Therefore, an emotional stimulus appears necessary for inferring emotion. Indeed, affective changes that either occur in the absence of a stimulus or last longer than a few minutes are commonly referred to as changes in mood.

**Converging Operations**

Although stimulus, emotional experience, behavior, and psychophysiology are important components of emotion, none is necessary per se to infer emotional processing. For example, expressive outputs in behavior or physiology might be voluntarily edited or concealed (display rules; Ekman, 1972), and for well-learned emotional responses, the peripheral physiological component might simply be short-circuited (Damasio, 1994; Harris & Katkin, 1975; Mandler, 1975). Also, observations of panic attacks without anxiety (Kushner & Beitzman, 1990) suggest that emotional processing might occur without any effects on emotional experience. Similarly, whether a stimulus is emotional often depends on appraisal processes that can be idiosyncratic for the individual (e.g., Ellsworth & Scherer, 2003). Also, if the eliciting stimulus is internal, such as a thought or a memory, it will be hard for an observer to measure. A possible solution to this problem is to adopt a convergent operations approach (Campbell & Fiske, 1959; Kihlstrom, Mulvaney, Tobias, & Tobis, 2000). According to such a perspective, emotion is inferred from the convergence of several indicators, and the inference is particularly persuasive if the indicators come from diverse domains (Campbell & Fiske, 1959).

**A Model of Unconscious Emotion**

Figure 5.1 shows an information-processing model of emotion to illustrate different approaches to the study of unconscious emotion. As shown, conscious and unconscious mechanisms evaluate and appraise the stimulus input. These emotional (core) mechanisms produce changes in emotional experience (feeling), behavior, and psychophysiology. Although the emotion is not directly observable, it is inferred from the properties of the stimulus, as well as from changes in emotional experience, behavior, and psychophysiology (shaded areas in figure). As indicated by the bidirectional arrows among experience, behavior, and psychophysiology, these different components might interact among each other. For example, as stipulated by psychophysiological theories of emotion (e.g., James, 1884), psychophysiological changes might affect emotional experience (for review, see Wiens, 2005).

Although figure 5.1 suggests that conscious and unconscious evaluator mechanisms receive equal input, the relative proportion of conscious and unconscious processing might vary for different emotions. For example, in fear, unconscious processing might dominate conscious processing. Also, although the figure suggests that conscious and unconscious evaluator mechanisms have separate effects on emotional output, they might share similar pathways on emotional output. However, whereas the depicted scenario (with separate pathways) allows conscious and unconscious mechanisms to have qualitatively different effects on emotional output, a model with identical pathways for conscious and unconscious mechanisms could not account for qualitatively different effects.

As further illustrated in Figure 5.1, unconscious emotion might be indicated in at least two ways. In a stimulus-focused approach, the input is manipulated so that conscious awareness of the stimulus is eliminated (as represented by X). Because participants are not consciously aware of the emotional input, any emotional responses are due to unconscious mechanisms. As such, these changes in experience, behavior, and psychophysiology would be evidence for unconscious emotional processing or simply unconscious emotion (e.g., Öhman, 1999). In contrast, in an experience-focused approach, the critical evidence for unconscious emotion is that participants show convincing signs of emotional pro-
processing without reporting any changes in their emotional experience (feeling). Thus, because it is irrelevant whether the stimulus input is processed consciously or unconsciously, emotional processing can be fully intact except for any effects on emotional experience (Kihlstrom et al., 2000). This approach is discussed in a recent review by Berridge and Winkielman (2003). In a study by Winkielman, Berridge, & Willbarger, 2005, thirsty and nonthirsty participants were shown masked emotional faces (happy, angry, or neutral) and were asked to indicate the gender of the neutral masking faces. Afterward, they were asked to rate their moods and were also given a pitcher of a fruit-flavored drink, which they could pour into a glass and consume. Results showed that although thirsty participants reported no changes in their emotional experience (mood), their consumption behavior was affected by the emotional faces. Participants who were shown masked happy faces poured and consumed more than participants who were shown masked neutral faces. In contrast, participants who were shown masked angry faces showed the opposite pattern. These data suggest that emotional responses (as inferred from the emotional stimuli and the drinking behavior) can occur without any effects on conscious experience (Winkielman & Berridge, 2004). Because it appears irrelevant for the experience-focused approach whether or not the stimulus input is processed outside of awareness, similar conclusions would be drawn for results obtained with nonmasked pictures. Although some research has employed this experience-focused definition of unconscious emotion, the stimulus-focused approach has been employed more commonly. These studies have mainly used backward masking to degrade visual input.

Backward Masking and the Dissociation Paradigm

The method of backward masking is considered to be the most promising method for studying unconscious processes (Holender, 1986). When a visual stimulus (target) is shown briefly and followed shortly by another visual stimulus (mask), people often report that they are consciously aware only of the mask, not of the preceding target. Thus the target is backward masked. For example, if a picture of an angry face is shown briefly and followed by a neutral face, participants might report being aware only of the neutral, not the angry face; that is, the angry face is masked. Aside from backward masking, other masking methods are available. For example, in sandwich masking, the target is both preceded and followed by masks, and in energy masking, a light flash serves as the mask. However, we do not consider these further because backward masking is used most commonly. Also, because these methods have been used for a similar purpose, the central issues in masking apply to all methods.

The masking method has been commonly used in the context of the dissociation paradigm. The basic rationale of this experimental design is as follows: In order to test whether awareness of the target pictures is a necessary condition for responding, awareness has to be completely eliminated. If masking parameters are manipulated so that participants are completely unaware of the masked target pictures (i.e., awareness is eliminated), any emotional responses to the masked pictures would have to be unconscious, as they would occur without any direct involvement of consciousness. In contrast, if conscious awareness of the masked pictures were necessary for their processing, then no responses could be observed in the absence of conscious awareness. Therefore, to rule out the possibility that responding could be potentially due to residual conscious processing, awareness has to be completely eliminated (as illustrated by the X in Figure 5.1). But this approach is conservative, as awareness could well be an epiphenomenon in many cases of emotional processing. That is, although awareness might accompany many responses, it might not play a causal role in these processes.

When the dissociation paradigm is used, it has to be demonstrated that participants are completely unaware of the masked pictures. Typically, unawareness is indexed by null
sensitivity of the measure of awareness. That is, if participants perform at chance levels on the awareness measure, they are considered unaware of the masked pictures. If these participants, however, respond better than chance on the emotion measure, the dissociation between the measures of awareness and emotion is interpreted to indicate that the emotional effects were unconscious, as they occurred without participants being aware of the masked pictures (Figure 5.1).

Example Study: Ohman and Soares (1994)

Ohman and Soares (1994) adopted the dissociation paradigm to study unconscious fear processes in participants afraid of either spiders or snakes. In this research, awareness was indexed by self-report and performance on a forced-choice classification task, and emotional processing was indexed by skin conductance and emotional ratings. In the first of two experiments, fearful and nonfearful participants were presented with masked pictures of spiders, snakes, flowers, and mushrooms at various stimulus-onset asynchronies (SOAs) between target and masking pictures. On each trial, participants responded if the masked target picture was a spider, snake, flower, or mushroom and also rated their confidence. Figure 5.2 shows results from fearful participants. As shown, classification performance and confidence ratings varied over SOA. A similar pattern was observed for nonfearful participants (not shown). Because at an SOA of 30 ms or less, fearful and nonfearful participants performed at chance levels in classifying the masked pictures, Ohman and Soares concluded that participants were unaware of the masked pictures at these masking parameters. Nonetheless, when the same masking parameters were used with other participants who reported themselves to be highly fearful of either spiders or snakes (1994, experiment 2), they showed greater skin conductance responses to masked pictures of feared animals (either spiders or snakes) than to pictures of nonfearred animals (either snakes or spiders) and flowers and mushrooms. These findings are shown in Figure 5.3. Also, when participants were presented with examples of masked pictures, they rated feared pictures as more negatively valenced and arousing than nonfearred pictures. In contrast, nonfearful participants did not differentiate among the pictures in their skin conductance responses and emotional ratings. Also, when asked to identify the content of the masked pictures, both fearful and nonfearful participants were generally unable to do so. Because the masking parameters apparently prevented participants from becoming aware of the masked pictures (as indexed by self-report and chance performance on the forced-choice classification task), these results suggest that participants can show fear (as indexed by skin conductance and emotional ratings) even when they are not consciously aware of the feared pictures.

The Ohman and Soares (1994) study illustrates the use of masking in the dissociation paradigm to study unconscious emotional processes. Although some readers might consider

Figure 5.2. Percent correct (top) and confidence ratings (bottom) for different pictures at various stimulus onset asynchronies (SOAs) for fearful participants in the Ohman and Soares (1994) study. Note that in the control condition (C), a masking picture served as the target. Adapted from Ohman and Soares, 1994.

Figure 5.3. Skin conductance responses to masked pictures for snake-fearful, spider-fearful, and nonfearful participants in the Ohman and Soares (1994) study. Adapted from Ohman and Soares, 1994.
the findings to be convincing data for unconscious emotions (fear), other readers might be more critical. In fact, there is an ongoing debate on a number of issues that are relevant to this type of research. However, before we delve into this discussion, we describe the main technical challenges in masking research. We have long experience with masking in our lab, and these comments and suggestions are intended to be helpful to other researchers who want to use masking in their labs. However, readers who want to go directly to the discussion of conceptual and methodological issues can skip the following section (and go to the section titled "Being aware of awareness, exhaustively and exclusively"). Also, Table 5.1 provides a summary of the main conceptual, methodological, and technical issues in masking, together with comments and suggestions.

The Nuts and Bolts of Masking

CRT Monitors

Masking requires that pictures be presented briefly and in short order. Also, valid masking requires that picture duration is stable over repeated trials. Because research suggests that small changes in picture duration can have strong effects on awareness (e.g., Esteves & Ohman, 1993), variability in picture duration over trials confounds changes in picture duration with changes in awareness. Indeed, this potentially confusing effect can seriously threaten the validity of masking results. For example, a performance index across trials is invalid if individual trials vary substantially in duration (and thus awareness). In fact, with substantial variability in picture duration and awareness, it appears hard to maintain perceptual processing at a particular performance level or threshold (e.g., unconscious perception). Hence, unless the reliability of masking set-up is demonstrated, the potential variability in picture durations over trials creates risks for confounded results (Wiens & Ohman, 2005). At first, reliable presentation of brief pictures might not seem to be a problem. For example, with common computer monitors (the bulky type), even film clips can be viewed without any flickering. In fact, because common monitors are cathode-ray tube (CRT) monitors, pictures can be presented for less than a few milliseconds. In a typical CRT monitor, a CRT beam moves rapidly across the screen and activates a thin phosphor layer that covers the entire screen. Typically, the beamer starts in one of the upper corners, continues horizontally to the end of the line, and then writes the next line below. When all lines are written, the beamer jumps back to the starting position. This whole cycle is called a refresh cycle. As an illustration, imagine that you try to write on a screen with a laser pointer. Because you can draw only a small part of the screen at a time, you have to move your pointer quickly across the screen to fill the screen. As part of a recent study, we conducted photodiode measurements to assess the accuracy of a common CRT monitor in presenting pictures at short durations (Wiens et al., 2004). In this study, a photodiode was placed in the middle of the monitor to detect luminance changes in response to picture presentations. The target picture was a white rectangle filling the entire screen. As shown in Figure 5.4, the pattern of luminance changes resembled a waveform. Each wave corresponded to a refresh cycle. Before the beamer reached the location of the photodiode (middle of the screen), luminance was zero; when the beamer swept over the location of the photodiode, luminance increased rapidly; and afterward, luminance dropped back to zero within a few milliseconds. However, luminance decreases occurred more gradually than luminance increases. This occurred because the phosphor remained activated even after the beamer had moved on.

Although our measurements showed that timing accuracy was excellent, the CRT technology has several limitations for masking purposes. First, interpicture intervals are rather long and can be manipulated only in large steps. Because it takes some time for the beamer to sweep over the whole screen (refresh cycle), the interval between two updates is determined by the time it takes for the beamer to complete a refresh cycle. This interval can be calculated from the refresh rate, which is the number of refresh cycles per second. For modern CRT monitors, refresh rates typically range from 60 to 160 Hz. Thus the interval between two consecutive screen updates (refresh cycles) varies from 16.7 ms (= 1/60 x 1,000) down to 6.3 ms (= 1/160 x 1,000), respectively. However, for most monitors, there is an inverse relationship between screen resolutions and supported refresh rates. That is, because the numbers of individual screen locations (i.e., pixels) increase at higher resolutions, supported refresh rates tend to decrease, as it takes longer to update all pixels. Therefore, before purchasing a monitor, we recommend checking which refresh rates are supported by particular screen resolutions. For example, although some monitors may allow refresh rates up to 160 Hz, their software drivers may be limited to 100 Hz. Further, although interpicture intervals of 6–17 ms might seem to be adequate SOAs between targets and masks, our own observations suggest that, depending on the type and size of the pictures, participants might report seeing the masked targets. Also, because common software does not allow one to change refresh rates during an experiment, the SOAs can be manipulated only in multiples of refresh rates. Thus, if the refresh rate is set to 60 Hz, the SOA can be manipulated only in intervals of 16.7 ms, from 16.7 to 33.4 to 50.1 ms, and so forth. From our experience, these changes in SOAs are too coarse, as they can result in dramatic changes in participants' perception of the masked targets.

A second limitation of CRT monitors is that picture duration is not continuous but a function of phosphor persistence and number of refresh cycles (Bridgeman, 1998). Pictures are not presented at once but are built up during a refresh cycle. Although this presentation is not ecologically


<table>
<thead>
<tr>
<th>Topic</th>
<th>Issues</th>
<th>Comments and Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conceptual/methodological</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stable picture presentation</td>
<td>For each condition, picture duration must be stable over trials, otherwise, variability in picture duration might confound changes in duration with changes in awareness.</td>
<td>Our research suggests that displays based on LCD/TFT technology do not fulfill this requirement whereas CRT and mechanical shutters do (Weins et al., 2004; Wiens &amp; Ohman, 2005).</td>
</tr>
<tr>
<td>Dissociation paradigm</td>
<td>Purpose is to eliminate awareness and test whether emotional responding remains. If so, this suggests that awareness is not necessary for emotional activation (cf. Holender, 1986).</td>
<td>A drawback is the requirement of null sensitivity on the awareness measure (see below). The study of qualitative differences and the psychophysical approach avoid this problem (see below).</td>
</tr>
<tr>
<td>Defining unawareness in terms of null sensitivity</td>
<td>Because unawareness is typically indexed by null sensitivity on the awareness measure, it depends on a statistical criterion and thus on power (e.g., alpha, number of trials).</td>
<td>The study of qualitative differences and the psychophysical approach avoid this problem (see below).</td>
</tr>
<tr>
<td>Defining a valid measure of awareness</td>
<td>A valid awareness measure has to capture all aspects of awareness (exhaustive) without capturing also unaware processes (exclusive; Merkle &amp; Reingold, 1998; Reingold &amp; Merkle, 1990).</td>
<td>Researchers have yet to agree on a valid measure of awareness. Whereas some favor objective measures (i.e., discrimination performance), others favor subjective measures (i.e., self-reported awareness; Wiens &amp; Ohman, 2002).</td>
</tr>
<tr>
<td>Measures assess thresholds</td>
<td>It is often assumed that an objective measure assesses the objective threshold. However, factors such as motivation, number of trials, and response bias can affect results. Hence, an objective measure might actually measure the subjective threshold (Merkle &amp; Daneman, 2000).</td>
<td>Because there is no generally accepted measure of awareness, research reports need to include sufficient detail about task and performance.</td>
</tr>
<tr>
<td>Qualitative differences</td>
<td>If observed effects differ qualitatively for two masking conditions, these findings support the notion of independent processes for unaware and aware processes and also provide evidence for the validity of the awareness measure (for a review, see Merkle &amp; Daneman, 2000).</td>
<td>Claims for qualitative differences need to be carefully examined.</td>
</tr>
<tr>
<td>Psychophysical approach</td>
<td>Determine the dose-response relationship between awareness and emotional processing.</td>
<td>Provides quantitative information about emotional processing at various levels of awareness. Advocates the use of comprehensive measures of awareness. Minimizes risk of missing critical masking parameters. Avoids proving null sensitivity of awareness measure (as for dissociation paradigm).</td>
</tr>
<tr>
<td><strong>Technical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRT monitor</td>
<td>Because CRT monitors write the screen on each refresh cycle, picture updates are limited by the refresh rate (e.g., 60 Hz = 16.7 ms), and the number of refresh cycles determines picture duration (Figure 5.4; Bridgeman, 1998; Wiens et al., 2004). Note that phosphor persistence in a single refresh cycle determines minimum picture duration.</td>
<td>CRT monitors exhibit excellent accuracy, and most experiment software allows synchronization with refresh cycles (Wiens et al., 2004). However, picture duration is limited to multiples of refresh cycles, and CRT monitors cannot be used in MRI because of interference.</td>
</tr>
<tr>
<td>TFT/LCD displays</td>
<td>Individual screen locations can be updated independently from each other. Because refresh cycles are not generated as for CRT monitors, experiment software has difficulties with synchronization. Also, picture presentation shows poor accuracy (Wiens et al., 2004).</td>
<td>Our results (Wiens et al., 2004) argue against the use of TFT/LCD displays. However, a notable feature is that pictures can be shown in steady state (i.e., no rewriting of the screen with each refresh cycle). This permits the use of LCD data projectors (instead of slide projectors) together with mechanical shutters in masking (see below).</td>
</tr>
<tr>
<td>Mechanical shutter</td>
<td>Mechanical shutters show excellent accuracy and can be controlled in milliseconds (independent from refresh cycles) (Wiens et al., 2004). Our research suggests that displays based on LCD/TFT technology do not fulfill this requirement, whereas CRT and mechanical shutters do (Wiens et al., 2004; Wiens &amp; Ohman, 2005).</td>
<td>Because data projectors (based on TFT/LCD technology) can display pictures in steady state, a shutter can be placed in front of the projector to control picture duration (Wiens &amp; Ohman, 2005). Therefore, we recommend the use of a mechanical shutter that is mounted in front of a data projector (for each target and masking picture; Figure 5.6).</td>
</tr>
</tbody>
</table>
valid, few vision researchers have discussed its implications (Krantz, 2000). However, even if this fact is ignored, it is important to keep in mind that picture duration is not the same as the duration of a refresh cycle. As illustrated in Figure 5.4, picture duration depends on phosphor persistence rather than on the refresh cycle. Therefore, even if a picture is nominally shown for only a single refresh cycle, its actual duration depends on phosphor persistence, which typically lasts 4 ms. Thus, whether a refresh cycle lasts 6.3 or 16.7 ms (at 160 or 60 Hz, respectively), the duration of a picture for a single refresh cycle would be about 4 ms. As further shown in Figure 5.4, actual picture duration is more difficult to determine if a picture is shown for several refresh cycles (Bridgeman, 1998).

**LCD/TFT Displays**

Although the CRT monitor has these limitations, its timing accuracy is outstanding compared with a modern flatbed monitor. This newer type of monitor is based on liquid crystal device (LCD) and thin-film transistor (TFT) technology. Simply stated, this type of monitor consists of a filter (LCD or TFT) that is placed in front of a light source. In contrast to CRT, this technology allows one to manipulate pixels in the filter individually so that it might be possible to update all pixels at once. Also, because the pixels can be held stable until they receive a new signal, picture duration can be continuous (this is illustrated in Figure 5.5C). However, although the LCD/TFT technology seems promising, our tests suggested that it might not be suitable for masking at this stage of technological advancement. That is, our photodiode measurements indicated that parameters such as onset latency, rise time, duration, and maximum luminance showed poor reliability for the panel, particularly so for short picture durations. Because small changes in picture parameters tend to have strong effects on perception (e.g., Esteves & Ohman, 1993) and thus may confound results, these data argue strongly against the use of LCD/TFT monitors in masking.

Similar results were obtained for a data projector with LCD technology. A data projector functions similarly to a flatbed monitor, except that the light source is stronger. Although Figure 5.5B illustrates that picture duration can be continuous, our photodiode measurements found that presentation parameters showed poor reliability. These results argue against the use of data projectors in masking.
Figure 5.5. Luminance changes over time for mechanical shutter (A), LCD data projector (B), and TFT panel (C) at picture durations of 8 (only for shutter), 12, 17, 24, 33, 47, 67, 82, 100, 106, 117, 141, 150, and 200 ms. Note that for projector and panel, durations at a refresh rate of 60 Hz are indicated by a dashed line. From Wiens et al., 2004.
Mechanical Shutters

Although these data support the use of CRT monitors, the drawbacks of CRT monitors (see preceding discussion) are reduced, if not eliminated, with mechanical shutters. As part of our study (Wiens et al., 2004), we also tested the timing accuracy of a mechanical shutter that was mounted on a data projector. A mechanical shutter resembles a metal screen that functions like the shutter in a camera; that is, a brief opening of the shutter controls exposure duration. In our lab, we mount the shutters in front of data projectors. Although mechanical shutters have traditionally been used together with slide projectors, we use data projectors. This is possible because pictures on a data projector are not rewritten as for CRT monitors (see Figure 5.5B and Figure 5.4). To test the accuracy of the shutter, a presentation screen was placed in front of the shutter, and the photodiode was fastened in the middle of the screen area that was illuminated by the data projector when the shutter was open. The data projector presented the target picture continuously, and the shutter controlled target durations. As shown in Figure 5.5A, the shutter showed outstanding precision. Further, because the duration of shutter opening is not limited by refresh rates at all, picture durations can be manipulated in steps of milliseconds. Therefore, the shutter permits excellent control over picture durations.

Recommendations

These features favor the use of mechanical shutters over CRT monitors. However, researchers might be concerned that the use of mechanical shutters requires a more demanding setup for masking than CRT monitors. On CRT monitors, it is possible to present target and masking pictures in short order, with a minimum SOA corresponding to a refresh cycle. In contrast, a setup with shutters requires a shutter and projector each for target and mask, as data and slide projectors do not allow one to change pictures within a few milliseconds. However, although hardware demands might be lower for CRT monitors than for shutters, the software demands for CRT monitors can cause difficulties of their own. The main problem is that the software needs to synchronize picture presentation with the refresh rate of the CRT monitor. That is, the software needs to know when a refresh cycle starts in order to update the screen only at the beginning of the new refresh cycle. If this is not done, pictures might be shown for either fewer or more refresh cycles than intended. Some available software synchronizes picture presentations with the refresh cycles of the CRT monitor. However, it appears that this software cannot control the exact onset of individual refresh cycles and is limited by particular refresh rates. In fact, even if it were possible to switch refresh rates within an experiment, our findings (see Figure 5.4) suggest that overall luminance levels might differ for different refresh rates, probably because there is less time for the beam to illuminate the phosphor at higher refresh rates. In contrast, software demands for the use of mechanical shutters are much simpler, as they require only a simple on/off signal (e.g., a TTL signal from the parallel port). Because shutters permit excellent control over onset and duration of pictures in steps of milliseconds, their use is recommended.

Figure 5.6 shows a schematic of a possible setup involving two mechanical shutters, two data projectors, and two computers (Wiens & Öhman, 2005). Both target and mask require their own data projectors and shutters to allow for immediate switching between target and mask. Figure 5.7 shows a picture of the setup in our lab with two casel VS25 shutters that are mounted in front of two data projectors. As further shown in Figure 5.6, a master computer displays the target picture on the target projector and sends a signal (e.g., TTL via parallel port) to the slave computer to display the mask. The master computer also controls target and mask shutter durations (e.g., TTL via parallel port). Because data projectors (unlike CRT monitors) present pictures in steady state, they can be used in combination with mechanical shutters that control picture duration. However, because data projectors need some time to reach steady state for each picture (e.g., Figure 5.5B), pictures ought to be shown only after steady state is reached. For example, in our lab shutters are opened only at least 200 ms after picture onset to guarantee steady state.

Taken together, our findings do not support the use of monitors and projectors based on LCD/TFT technology in masking. In contrast, both CRT monitors and mechanical shutters showed excellent timing accuracy. However, mechanical shutters permit a more ecologically valid method of picture presentation than CRT monitors. Also, they permit exact control over picture onset and duration in steps of milliseconds rather than in steps of refresh cycles. Because our own experiences indicate that small changes in masking parameters can have strong effects on participants' perception of the masked pictures, the advantage of greater control over stimulus parameters recommends the use of mechanical shutters in masking. But, because our results might not apply to all available and future displays (e.g., plasma displays, displays with digital light processing technology, LCD goggles), we encourage researchers to test their equipment for accuracy. Indeed, Mollon and Polden (1978) tested tachistoscopes and demonstrated that accuracy was generally poorer than claimed by manufacturers and assumed by researchers. Because our observations confirm that the validity of presentation parameters cannot be taken at face value, we recommend that researchers provide evidence for the accuracy of their display equipment in their publications. Photodiode measurements might be ideal for that purpose (for details, see Wiens et al., 2004). Although evidence for accuracy is particularly important for displays based on LCD/TFT technology, even studies that use shutters and CRT monitors should provide this evidence. For example, studies with CRT monitors should include information regard-
ing refresh rate and ensure that picture presentation was synchronized with the refresh cycle.

**Being Aware of Awareness, Exhaustively and Exclusively**

Aside from these technical aspects of masking, there are a number of conceptual issues that continue to be debated in the literature. The extensive use of the dissociation paradigm (e.g., Ohman & Soares, 1994) might be due to the fact that it was advocated in a widely cited review (Holender, 1986). Unfortunately, it has a number of problems, most of which stem from two basic requirements: to find a valid measure of awareness and to demonstrate that this measure shows null sensitivity (i.e., that participants perform at chance levels).

**Exhaustiveness: Measuring All Aspects of Consciousness**

As discussed by Reingold and Merkle (1990), a valid measure of awareness ought to be exhaustive in that it measures all conscious aspects of awareness. The distinction between a nonexhaustive and an exhaustive measure is illustrated in Figure 5.8. In this figure, solid lines represent conscious input, and dashed lines represent unconscious input. Further,

![Figure 5.6. Schematic of the masking setup used in our lab.](image)

![Figure 5.7. Picture of the masking setup used in our lab with two VS25 shutters (Vincent Associates, Rochester, New York) that are mounted in front of two data projectors. Because our system needs to be mobile, clamps are used to hold the shutters in place.](image)
thick lines highlight input that is captured by the awareness measure. As shown in Figure 5.8A, if a measure is not exhaustive, it does not capture all aspects of conscious input. Hence, any purported unconscious effects might be due to the residual conscious input that is not captured by the measure (i.e., the thin lines in Figure 5.8A). In contrast, an exhaustive measure (Figure 5.8B) measures all aspects of conscious processing.

For example, Lovibond and Shanks (2002) criticized the awareness measure used by Ohman and Soares (1994), claiming that this forced-choice classification task is not sensitive enough to index awareness. That is, participants might be able to discriminate among the masked pictures without being able to identify (i.e., verbally label) them as spiders, snakes, flowers, and mushrooms. Accordingly, participants might have been aware of curvy features of the masked snakes without being able to identify these pictures as snakes. This argument implies that the classification task in the Ohman and Soares (1994) study was not an exhaustive measure of awareness. Thus, because Lovibond and Shanks (2002) argued that the classification task was invalid, the findings in the Ohman and Soares (1994) study cannot be taken as evidence that participants responded to feared masked pictures of which they were unaware.

Exclusiveness: Measuring Only Consciousness

Another central requirement for a valid measure of awareness is that it ought to be an exclusive index of only conscious, not unconscious, processes (Merikle & Reingold, 1998). Therefore, if unconscious, as well as conscious, processes affect a measure, it would violate this requirement, as it would not be an exclusive index of only conscious processes but would also be affected by unconscious processes. This is illustrated in Figure 5.8C. Although a measure might be exhaustive, it might not be exclusive. If so, this measure would not only be sensitive to all aspects of conscious processes but would also be partly sensitive to unconscious processes. Thus it would not allow one to determine whether performance is due to conscious or unconscious processes or both. Therefore, a valid measure has to be both exclusive and exhaustive; in that it measures all conscious processes but no unconscious processes.

For example, it is unclear whether the awareness measure proposed by Lovibond and Shanks (2002) fulfills the requirement of exclusiveness (Wiens & Ohman, 2002). If participants’ ability to discriminate among the masked pictures without being able to identify them is not an exclusive measure of conscious processes, discrimination could be due to unconscious and/or conscious processes (Figure 5.8C). If so, the awareness measure suggested by Lovibond and Shanks (2002) would be invalid.

Subjective and Objective Thresholds

The debate about the validity of different measures of awareness has a long history (Eriksen, 1960). In general, this debate centers on a distinction between awareness measures that index either objective or subjective thresholds (Cheesman & Merikle, 1984, 1986). Originally, Cheesman and Merikle (1984) defined the subjective threshold as the “level at which subjects claim not to be able to discriminate perceptual information at better than chance level” and the objective threshold as the “level at which perceptual information is actually discriminated at a chance level” (p. 391). During the past decade, however, these terms have been used more generally to distinguish between people’s (subjective) self-reported or claimed awareness and their actual (objective) discrimination performance (e.g., Merikle & Daneman, 2000). Therefore, in the present context, people would be subjectively unaware of the masked pictures if they reported that they could not tell what was depicted in the pictures. In contrast, people would be objectively unaware of the masked pictures if they could not discriminate among the masked pictures. Based on this dis-

Figure 5.8. Schematic of effects of exhaustiveness and exclusiveness on measuring awareness. Solid lines represent conscious input, and dashed lines show unconscious input. Further, thick lines highlight the input that is captured by the awareness measure. Note that a measure that is not an exhaustive measure (A) is not sensitive to all aspects of conscious processes, whereas a measure that is not exclusive (C) is sensitive also to unconscious processes. An ideal measure of awareness is both exhaustive and exclusive (B).
tinction, tasks that measure participants' ability to discriminate among the masked pictures are commonly called objective measures, whereas tasks that measure participants' self-reported awareness are called subjective measures. For example, the classification task in the Ohman and Soares (1994) study was an objective measure, as participants' ability to recognize masked spiders, snakes, flowers, and mushrooms was assessed with a forced-choice classification task in which participants chose among all possible picture categories. In contrast, presenting participants with masked pictures and asking them to report what was depicted would be a subjective measure of awareness, as participants' self-reported awareness is measured. Research suggests that objective and subjective measures yield different results; objective measures typically require stronger masking (e.g., shorter SOA) than subjective measures for participants to be classified as unaware (Cheesman & Merkle, 1984, 1986).

Objective measures have generally been favored over subjective measures (Holender, 1986). The most important reason is that objective measures do not leave the response criterion to the participant (Eriksen, 1960). As captured in signal detection theory (Macmillan & Creelman, 1991), participants' responses on a discrimination task are affected by their discrimination abilities (e.g., indexed by d'), as well as their response biases (e.g., indexed by beta; e.g., Wiens, Emmertich, & Katkin, 1997). Response bias reflects participants' predisposition to respond that the target was present. Whereas objective measures often allow one to separate discrimination ability from response bias, subjective measures might not allow this. To illustrate with a simple example, if the subjective measure described earlier is used, participants might differ in their willingness to speculate on what was depicted in the masked pictures. Because the conservative participants would perform less well on a subjective awareness measure than the liberal participants, it would seem that conservative participants are less aware of the masked pictures than liberal participants. Thus, even if conservative and liberal participants do not differ in their discrimination abilities, differences in response bias would result in confounded subjective measures of awareness.

Several factors can affect response bias. Whereas one common factor is demand characteristics (Eriksen, 1960), Merkle (1992) pointed out that subjective measures also reflect participants' notions of the value of their own awareness. For example, some participants might report that they are aware of a snake picture if they can detect something clearly, whereas other participants might report that they are aware of a snake picture only if they can clearly identify all aspects of the snake. However, because the experimenter has little control over the participant's criteria and response biases, "we are in fact using as many different criteria of awareness as we have experimental subjects" (Eriksen, 1960, pp. 291-293). Also, because research has shown that participants tend to underestimate their performance in difficult perception tasks (Björkman, Juslin, & Winman, 1993, Cheesman & Merkle, 1984, 1986), participants in masking studies might be biased toward being less willing to report that they were aware of masked pictures. As a result, subjective measures might well underestimate participants' awareness of the masked pictures. Consistent with this expectation, subjective measures of awareness tend to be less sensitive than objective measures (e.g., Cheesman & Merkle, 1984). In sum, because objective measures typically allow one to separate discrimination ability from response biases, they can control for the confounding influence of performance underestimation.

The Subjective Nature of Awareness: Noticing Versus Perceiving

Although objective measures of awareness allow one to assess participants' discrimination ability objectively, this feature is also their greatest drawback. In contrast to objective measures, objective measures transfer the responsibility of deciding about awareness to the experimenter. However, as elaborated by Bowers (1984), awareness is basically of a subjective nature, a fact that is completely ignored by objective measures of awareness. Because awareness is a process that is closer to the process of "noticing" than "perceiving," a valid measure of awareness ought to capture what participants notice rather than what they can perceive. Therefore, the critical issue is whether self-reported awareness or objectively measured discrimination ability is a better index of awareness. Adherents of objective measures rely on discrimination ability (e.g., an experimental psychologist's perspective) and dismiss participants' self-reports as confounded by response biases. However, because objective measures index only participants' ability to discriminate (i.e., to perceive instead of notice), they ignore the inherently subjective nature of awareness and thus cannot be viewed as valid measures of awareness per se (Wiens & Ohman, 2002). Therefore, findings that objective measures tend to be more sensitive than subjective measures might be due to their violation of exclusiveness rather than to response biases.

Are Unaware Processes Logically Impossible?

In fact, whereas proponents of objective measures of awareness generally accept that performance at chance levels indicates that people are unaware of masked pictures, they infer that performance better than chance demonstrates necessarily that people are actually aware of the masked pictures. This reasoning has been challenged by studies of brain-damaged patients. In these studies, objective measures were used to demonstrate that people can perform a task although they are unaware of the target stimuli, as indexed by subjective measures. The most famous example is blindsight. As reported by Weiskrantz (2000), patients with damage to the primary visual cortex report that they are completely unaware of visual stimuli presented to their damaged visual field.
Further, when instructed to respond to these stimuli, patients report that they are purely guessing and need to be urged to respond anyway. Nonetheless, the patients show evidence that they can discriminate among the stimuli. Thus the patients exhibit blindsight in that they can discriminate stimuli in their damaged visual field although they report that they are unaware of the stimuli. Because objective measures (i.e., discrimination tasks) were used to index performance outside of awareness, these data challenge the notion that above-chance performance on objective measures necessarily shows that people are aware of the target stimuli. In fact, the notion that greater-than-chance performance demonstrates awareness implies that blindsight, as well as similar perceptual phenomena in unilateral neglect and prosopagnosia, does not exist, as greater-than-chance performance on objective measures would necessarily indicate that patients are aware of the target stimuli. Accordingly, if this reasoning were taken to the extreme, then any kind of discriminative response to masked pictures would necessarily indicate awareness. If so, it would be logically impossible to demonstrate performance without awareness (Bowers, 1984), as any form of discriminative responding would, by definition, indicate that participants were aware of the masked pictures (Wiens & Ohman, 2002).

**Awareness as a Statistical Criterion**

Aside from requiring a valid measure of awareness, the dissociation paradigm also requires that this measure show null sensitivity. In fact, even if researchers can agree on how to measure awareness, the requirement for demonstrating null sensitivity represents a problem of its own. If unawareness were defined in terms of an absolute performance level, then unawareness would not depend on any other factors aside from the absolute performance level. To illustrate, imagine a task that requires participants to recognize masked pictures of spiders and snakes, in which half of the pictures are spiders and half are snakes. On each trial, either a masked spider or snake is presented, and participants respond by forced choice whether the masked picture was a spider or a snake. If unawareness were defined in terms of an absolute performance level of 50%, which would be expected if participants responded randomly, then unawareness would not depend on any other factors. Accordingly, individual participants would be unaware if their performance levels were at 50%.

**Individual Testing: You Must Be Aware ($p < .05$)?**

However, because null sensitivity is commonly defined in terms of chance performance, unawareness is based on a statistical criterion. Therefore, in order to determine whether an individual participant performed at chance (and was unaware of the masked pictures), awareness must depend not only on the absolute performance level but also on the significance level (i.e., alpha) and the number of trials. As a result, awareness depends on the statistical power of the awareness test. Regarding effects of alpha, imagine that the forced-choice classification task of masked spiders and snakes consisted of 20 trials. Then a one-tailed alpha of .20 would correspond to a performance level exceeding 60% correct (>12 trials), whereas a one-tailed alpha level of .10 would correspond to a performance level exceeding 65% correct (>13 trials). That is, at an alpha of .20, individual participants would be unaware if they identified 60% or fewer of the masked spiders and snakes, whereas at an alpha of .10, participants would be unaware at up to 65% correct identifications. Thus the choice of alpha affects whether or not participants are classified as unaware. Further, although it might seem that alpha levels of .10 and .20 are odd and that the traditional alpha of .05 would do, the choice of the standard alpha of .05 might not be appropriate. Because most research is interested in rejecting the null hypothesis, an alpha of .05 allows for a low probability of only 5% of rejecting the null hypothesis even though it is true (i.e., Type I error). However, although a low alpha reduces the risk for a Type I error, it increases the risk for a Type II error. That is, although an alpha level of .05 reduces the risk of falsely concluding that participants who perform better than chance are aware of the masked pictures even though they are not (i.e., Type I error), it increases the risk for falsely concluding that participants who do not perform better than chance are unaware of the masked pictures even though they are (i.e., Type II error). In our example with 20 trials, participants would have to identify more than 70% correct to perform significantly above chance at a one-tailed alpha of .05. Thus participants who perform at or below 70% correct would be unaware of the masked spiders and snakes. However, the validity of this conclusion might be questionable, particularly so for performance levels at or close to 70% correct. Therefore, if the purpose of the research is to show that participants are unaware of the masked pictures, a big alpha level (at least .10) ought to be selected to reduce the risk for a Type II error. However, because there is no standard alpha level in masking research, researchers have set alpha levels arbitrarily.

Also, issues of directionality need to be considered when alpha is set. Because alpha can be one-tailed or two-tailed, it can be determined if individual participants perform worse than chance, as well as better than chance. In terms of the example, participants might perform significantly above chance in recognizing spiders and snakes but might, alternatively, perform significantly below chance in recognizing spiders and snakes. In the former case, participants correctly label spider pictures as spiders and snake pictures as snakes, whereas in the latter case, they reverse label spiders and snakes. Performance below chance levels is hard to interpret, as it suggests that participants could apparently discriminate among the masked pictures. Nonetheless, because such reverse labeling can happen, it is important to clarify whether the choice of alpha is one-tailed or two-tailed.
Further, when unawareness for individual participants is determined based on a statistical criterion, it depends not only on the alpha level but also on the number of trials. In the above example with 20 trials, individual participants performed better than chance at a one-tailed alpha of .10 if they correctly identified more than 65% (> 13 trials). However, if the task were based on 40 trials, individual participants would already perform better than chance (at one-tailed alpha = .10) if they correctly identified more than 60% (> 24 trials). In short, as proportionally fewer trials are sufficient to obtain a statistically reliable estimate, an increase in number of trials increases statistical power.

**Group Testing: Everybody Is Aware!**

Whereas the preceding examples referred to the situation of determining whether individual participants performed better than chance, many studies tested whether a group of participants performed better than chance. These studies typically reported that participants in the study were unaware of the masked stimuli, because they did not perform better than chance as a group. However, it is questionable that a global significance test is sufficient to demonstrate unawareness for the whole sample. For example, if participants vary substantially in their performance (i.e., high variance), it is unlikely that a global performance test will turn out significant. If so, the sample could contain a number of participants who performed quite well. Because this would be consistent with the idea that some participants were aware of the masked pictures, it would be erroneous to conclude, based on a global significance test, that all participants were unaware of the masked pictures.

Also, similar problems to those discussed for significance testing of individual participants apply to significance testing of a group of participants. Accordingly, the choice of both size and directionality of alpha are important, as well as statistical power. Whereas testing of individual participants varies with the number of trials in the awareness measure, testing of a group varies with the number of participants. For example, if a group of participants went through the spider/snake classification task and obtained overall 52.7% correct (SD = 5.2), mean group performance would not differ significantly from chance performance (50%) with a sample size of 10, t(9) = 1.63, p = .137. In contrast, with a sample size of 20, group performance would differ from chance, t(19) = 2.37, p = .027. In this example, the additional participants in the larger group obtained identical scores to the first group. Therefore, if unawareness is determined for the whole sample, then these findings would lead to the nonsensical conclusion that participants become aware as sample sizes increase. To conclude, problems in defining unawareness based on a statistical criterion are apparent at a group level, as well as an individual level, of statistical testing.

The difficulties in using a statistical criterion in defining unawareness affect mostly objective rather than subjective measures of awareness, as objective measures rely on behavioral performance data rather than on self-report. As subjective measures assess participants' self-reported awareness, statistical testing is of minor importance in determining whether participants are unaware of the masked pictures. In contrast, as objective measures assess participants' discrimination performance, statistical testing is commonly used to decide whether participants are unaware of the masked pictures. Taken together with the conceptual problem that objective measures are insensitive to the subjective nature of awareness, these issues question the validity of indexing awareness with objective measures.

**Measures Versus Thresholds**

*When an Objective Measure Indexes the Subjective Threshold*

In fact, it is commonly believed that the use of objective measures ensures that the objective threshold is assessed; however, this assumption is challenged by several problems. First, even if awareness is measured with an objective task, it is possible that the performance measure used might be confounded by response bias. As such, it would not be a pure (objective) index of participants' discrimination ability but might represent a measure of participants' subjective awareness. For example, although detection tasks are objective measures, the commonly used performance measure of percent correct is affected by response bias, as well as discrimination ability. That is, if participants are instructed to detect whether or not a target was shown, percent correct will be affected by their predisposition to respond that a target was shown. In fact, due to changes in response bias, a participant with stable discrimination ability might score anywhere from 50–95% correct on detection tasks (Azzopardi & Cowey, 1997).

Second, if a statistical criterion is used to define unawareness, the classification of participants as aware or unaware depends on the statistical power of the objective task. For example, if the objective task consists of only a few trials, participants could perform almost perfectly and still be classified as unaware (Merkile, 1982). To illustrate, if the example task involving masked spiders and snakes (discussed earlier) consisted of 10 trials, individual participants could identify correctly 70% of the pictures without performing significantly better than chance (at one-tailed alpha = .10). However, the conclusion that the participants were unaware of the masked pictures at the objective threshold appears to have questionable validity.

Third, when masking parameters are chosen so that the masked pictures are barely, if at all, visible, participants might have no motivation to perform a discrimination task (Duncan, 1985; Merkile, 1992). Because participants would report being unaware of the masked pictures, a forced-choice
classification task that requires participants to indicate whether a masked picture was a spider or a snake might be perceived as meaningless, as none of the pictures are noticed. If so, participants might not see a point in trying to classify spiders and snakes and, therefore, might push buttons randomly. As a result, they would perform at chance. Thus the objective measure (i.e., identification task) would not necessarily index the objective threshold but might reflect the subjective threshold (i.e., participants’ self-reported unawareness). Our own observations from pilot studies support this argument. When pictures were masked well and participants were instructed to indicate whether a masked picture was a spider or a snake, some participants stopped responding because they thought that the projector was broken, as they did not notice any spiders and snakes. Therefore, in our experiments in which we attempted to assess the objective threshold, we explained to participants that it might appear to them that no target pictures are shown at all but that they should attend to all pictures and decide whether a spider or snake was presented. Thus participants’ subjective sense of guessing on the classification task did not undermine their motivation to perform well on this task. However, this discussion highlights the issue of whether or not the use of objective measures guarantees that the objective threshold is assessed. For example, Merkle and Daneman (2000) suggested that factors such as an insufficient number of trials on the objective measure, as well as insufficient motivation of the participants, could explain why many reports of unconscious processes have been comparable in studies in which unawareness was assessed with either an objective measure (purportedly indexing the objective threshold) or a subjective measure.

To conclude, the use of an objective task does not guarantee that the objective threshold is assessed. However, if a performance measure is used that is unaffected by response biases (e.g., d’ instead of percent correct), if the task has sufficient statistical power (e.g., sufficient number of trials, large alpha), and if the participants are motivated to perform the task, then the finding that participants performed at chance would support the conclusion that participants were unaware at the objective threshold. However, in order to decide about the validity of claims regarding unawareness, these matters ought to be addressed in every research report. For example, one suggestion is to include confidence intervals, as these depend on statistical power. However, because there are no generally accepted criteria (e.g., size of alpha), any conclusion regarding awareness might be questioned by researchers with different criteria.

**Measuring Thresholds**

Although we believe that the concepts of subjective and objective thresholds are helpful, they provide only rough guidelines in distinguishing different levels of awareness. In general, the subjective threshold refers to whether participants notice the masked pictures, whereas the objective threshold refers to whether participants can discriminate the masked pictures on a behavioral task. However, because the definitions are rather vague, it is unclear which specific measures ought to be used to capture these thresholds. In fact, Cheesman and Merkle (1986) measured thresholds in terms of performance on either a detection task (experiment 2) or an identification task (experiment 3). That is, performance was indexed by participants’ subjective or objective ability to detect whether or not a word was presented (detection task) or which word was presented (discrimination task). However, because research suggests that these tasks can yield different results (e.g., Haase, Theios, & Jennison, 1999), it is unclear whether these two tasks measure the same process. To give another example, the task used by Ohman and Soares (1994) to measure participants’ ability to classify masked spiders, snakes, flowers, and mushrooms appears to provide an index of the objective threshold. However, as Lovbod and Shank (2002) argued, participants might be able to discriminate among the masked pictures without being able to verbally label the masked pictures. Because both measures assess different kinds of discrimination abilities, it is unclear which measure is the correct one to index the objective threshold. Similarly, if participants are shown masked spiders and snakes, but they report that they did not see any spiders or snakes, it might be concluded that they were unaware (subjective threshold). However, if participants also reported that they noticed that masked pictures were shown even though they could not tell whether the pictures were spiders or snakes, these findings would suggest that participants did have some subjective awareness of the masked pictures (similar to Weiskrantz’s [2000] notion of type 2 blindsight).

Therefore, it is unclear whether it can be concluded that participants were unaware at the subjective threshold. So, unless one postulates that there are as many thresholds as there are tasks, conclusions regarding whether the subjective or objective threshold is assessed depend on the specific task that was used to index awareness. However, this decision might be particularly difficult for tasks that combine subjective with objective features. For example, Kunimoto, Miller, and Pashler (2001) introduced a measure that combines (objective) discrimination ability with (subjective) confidence. On each trial, participants choose a response and indicate their confidence (high or low). Then a signal-detection analysis is performed in which correct responses with high confidence (hits) are contrasted with incorrect responses with high confidence (false alarms). Because participants’ predisposition to indicate high confidence (response bias) affects both hits and false alarms, participants’ ability to discriminate correct responses from incorrect responses in terms of confidence can be assessed independent of response bias. Taken together, these points indicate that concepts of measurement and threshold need to be distinguished to avoid potential misunderstandings in the discussion of measuring...
awareness. Thus they illustrate the need to distinguish between empirical measures and theoretical constructs, a distinction that has often been blurred in the study of unconscious processes (Reingold & Merkile, 1990).

When to Measure Awareness

Because masking research on emotional responses that uses the dissociation paradigm includes two measures, one to index awareness and the other to index emotional processes, a final issue is when these measures ought to be taken. We recommend that both measures be taken in the same participants (i.e., within-subjects design). Because awareness is best tested at the level of individual participants, it is advisable to measure awareness of all participants in the study. However, if factors such as time restrictions do not permit obtaining both measures from the same participants, performance data from other participants are helpful in characterizing awareness for the participants in the study. For example, before their actual experiment, Ohman and Soares (1994) conducted an experiment in which they found that participants showed little, if any, evidence that they could classify masked spiders, snakes, flowers, and mushrooms at the particular masking parameters. These findings supported the notion that the masking parameters were generally effective in preventing participants from recognizing the masked pictures.

Further, when awareness and emotional processes are assessed in a within-subjects design, both measures can be collected concurrently or separately. In our research, we have used both approaches. For example, in the Wiers, Katzkin, and Ohman (2003) study, a separate forced-choice classification task was used, whereas in the Ohman and Soares (1998, experiment 2) study, some participants completed a classification task while emotional processes (i.e., skin conductance) were monitored. If awareness is assessed in a separate task, it is best measured in the same session as the emotional measures, as performance might not remain constant between sessions (e.g., Wollard, Marchak, & Hughes, 1988). However, although a concurrent measure of awareness might be most sensitive in indexing awareness, it is likely that the measuring process affects the variables of interest. Therefore, emotional processing might depend on the task that participants perform during the experiment. For example, when participants were shown emotional faces and had to decide whether one was afraid or unfamiliar, they showed amygdala activation only when they matched the faces to other faces, not when they verbally labeled the expressions (Harri, Bookheimer, & Mazzotta, 2000). Because concurrent and separate measures of awareness represent different experimental contexts that might engage different processes (e.g., passive viewing versus active search), this discussion suggests that the timing of measuring awareness needs to be carefully considered.

Measuring Emotions

Clearly, the main issues with masking research have to do with the assessment of awareness. However, because the purpose of using masking in emotion studies is to study emotional processes that occur outside of awareness, emotional measures need to be included as dependent variables. In our own research, we have mostly used psychophysiological measures of emotion with an emphasis on sweat gland activity. The study by Ohman and Soares (1994) is a typical example (see Figure 5.3). As described earlier, results showed that participants who were afraid of spiders (but not of snakes) showed greater skin conductance responses (SCRs) to masked spiders than to masked snakes, flowers, and mushrooms. Similarly, participants afraid of snakes (but not spiders) showed greater SCRs to masked snakes than to other masked pictures. In contrast, participants who were unafraid of spiders and snakes did not show elevated SCRs to these masked pictures. Because these data are consistent with reports that people with phobias show elevated SCRs to fear-eliciting pictures (Globisch, Hamm, Esteves, & Ohman, 1999), the findings by Ohman and Soares (1994) suggest that fear was elicited unconsciously by the masked spiders and snakes. However, research has found that people show elevated SCRs to arousing stimuli, whether negative or positive (for a review, see Bradley et al., 2001). For example, when people rated pictures on dimensions of valence and arousal, they showed elevated SCRs to pictures that they rated as unpleasant and arousing (e.g., accidents, mutilations), as well as to pictures that they rated as pleasant and arousing (e.g., nudes). Thus elevated SCRs were not specific to negative emotions but depended on how arousing the pictures were. Therefore, findings of elevated SCRs to masked spiders and snakes (Ohman & Soares, 1994) might reflect processes associated with the arousing properties of the spiders and snakes, rather than negative valence. Further, it is possible that such responses might represent attentional, as well as emotional, processes (Ohman & Wiens, 2003). Although findings of amygdala activation to nonmasked and masked phobic pictures in people with spider and snake phobias (Carlsson et al., 2004) provide evidence for emotion, this example illustrates that researchers need to be aware of the limitations of their emotional measures.

Masking: Too Many Problems?

At this point, critical readers of this chapter might conclude that the masking method has too many problems to be useful in studying unconscious emotional processes. The technical demands seem quite tedious. Also, given that there is no generally accepted measure of awareness, the choice of an awareness measure seems to be arbitrary. Further, when the dissociation paradigm is used, unawareness is commonly demonstrated with null sensitivity. However, because null
sensitivity is typically defined based on a statistical criterion, it depends on a number of factors (e.g., alpha, number of trials) that also seem to be arbitrarily chosen by researchers. In fact, whether or not significant results are obtained, critics with different criteria will probably challenge the data. If null findings are obtained, the study might never get published. With this "file-drawer problem," the generalizability of significant findings will be overestimated. Also, as discussed by Eriksen (1960), dissociations between two measures might occur due to chance. Thus the dissociation paradigm does not rule out the possibility that, even though the awareness measure shows null sensitivity, significant findings on an emotional measure might be due to chance (Type I error). Indeed, because the dissociation paradigm combines two statistical tests (i.e., null sensitivity for awareness measure and significance for the emotional measure), the actual probability of a Type I error is not readily apparent. Null findings are also problematic because there is no manipulation check for the perceptual processing of masked pictures. That is, although participants should be unaware of the masked pictures, they should be able to process them perceptually. Otherwise, no effects at all would be expected. For example, if masked pictures are shown for 1 nanosec, they would probably not be processed at all, and neither nonemotional perceptual nor emotional processes would be expected. Therefore, without a manipulation check for nonemotional perceptual processing to masked pictures, null findings for emotional processes are hard to interpret.

**Two Potential Solutions: Qualitative Differences and the Psychophysical Approach**

Although these problems might appear insurmountable, there are at least two solutions to these issues: demonstrating qualitative differences and taking a psychophysical approach. Merkle and his colleagues have developed and applied the approach of demonstrating qualitative differences to memory and perception (for review, see Merkle & Daneman, 2000). The central aspect of their approach is that obtaining effects that differ qualitatively for conscious and unconscious processes validates the distinction between these processes. That is, results that show that conscious processes differ in direction (i.e., qualitatively) and not just in size from unconscious processes make the distinction between both processes valid and interesting. Also, if a measure of awareness can index these qualitative differences between unconscious and conscious processes, it receives support as a valid measure of awareness.

Although it is challenging to find experimental contexts in which conscious and unconscious processes would be postulated to differ qualitatively, Merkle and his colleagues have identified several examples in independent areas of research (Merkle & Daneman, 2000). For example, in a study by Merkle and Cheesman (1987), participants were presented with a modified version of the Stroop task. In a typical Stroop task, participants are shown color words (e.g., green, red) that are printed in color (e.g., green, red). When the color words are printed in matching colors (e.g., green in green ink, red in red ink), participants have no difficulties in naming the ink colors. However, when the color words are printed in nonmatching colors (e.g., green in red ink, red in green ink), participants performance in naming the ink colors is impaired (Stroop effect), as it is difficult to ignore the semantic content. In the Merkle and Cheesman (1987) study, only the words green and red and the corresponding ink colors were used. On each trial, a word was shown, followed by a color patch. Participants were instructed to name the color of the patch as quickly as possible. In addition, they were informed that on 80% of the trials, word and color would not match, but on the remaining 20% of the trials, word and color would match. Thus, when participants saw a color word, they could anticipate that, for the majority of trials, the color of the patch would not match the color word (e.g., if they saw green, the color of the patch would most likely be red). To study unconscious and conscious processes, the color words were masked at either short or long SOAs. Merkle and Cheesman (1987) argued that when participants can consciously perceive the color words (at long SOA), they can use the information from the color words intentionally in anticipating the nonmatching ink. That is, when participants are aware of the words (at long SOA), they can use the relative frequency of matching and nonmatching events to improve their performance. Hence, participants should respond faster on nonmatching trials and more slowly on matching trials. In contrast, when participants are not aware of the words (at short SOA), they cannot use the information from the color words to counteract automatic processes. Hence, participants should respond faster on matching trials and more slowly on nonmatching trials. Thus participants would show a typical Stroop effect when words are masked (short SOA) but a reverse Stroop effect when words are not masked (long SOA). Consistent with these predictions, results showed qualitatively different effects for masked and nonmasked color words. When words were masked, reaction times were faster on matching than on nonmatching trials. In contrast, when words were not masked, reaction times were faster on nonmatching than matching trials. These results have been replicated (Daza, Orrelli, & Fox, 2002; Merkle & Joordeens, 1997). Because effects on reaction time were in the opposite direction for masked and nonmasked words, these findings provide evidence for qualitatively different processes. Thus they validate the distinction between conscious and unconscious processes. Because research further suggests that participants were unaware of the masked words at the subjective threshold (Cheesman & Merkle, 1984), the findings of qualitative differences suggest that the subjective threshold is a valid index of awareness.
Qualitative Differences

The approach of demonstrating qualitative differences could be useful in studying emotional conscious and unconscious processes. Unfortunately, to our knowledge, there is no study that has explicitly used this approach. However, the results of a study by Morris, Ohman, and Dolan (1998, 1999) might qualify as an example of qualitative differences (Wiens & Öhman, 2002). In this study, participants were fear conditioned to two nonmasked angry faces. That is, for each participant, it was randomly determined which of two angry faces was paired with a loud noise. At the beginning of the experiment, participants were shown both angry faces and were instructed that they would perform a detection task during the experiment. On each trial, if they saw either angry face, however fleetingly, they should push one button, and if they did not see either angry face, they should push another button. After conditioning, participants were shown masked, as well as nonmasked, examples of the angry faces. When angry faces were masked, neutral faces served as masks, and when angry faces were not masked, the angry faces were shown after the neutral faces. Results for the detection task showed that participants indicated that they saw all of the nonmasked angry faces but none of the masked angry faces. Further, results from positron emission tomography (PET) brain imaging indicated that, after conditioning, differential amygdala activation occurred for masked and nonmasked presentations of the angry faces. In nonmasked presentations, the left amygdala was activated, whereas in masked presentations, the right amygdala was activated. Because this pattern of results yielded a significant interaction between masking and laterality, the findings suggest qualitatively different effects for masked and nonmasked presentations on amygdala activation. These data replicate and extend prior research that demonstrated a critical role of the amygdala in conditioned fear (LeDoux, 1996).

Also, because the qualitative differences were indexed by the awareness measure used by Morris et al. (1998), these findings support the validity of the awareness measure. Thus these data challenge Lovibond and Shankes's (2002) criticism of this type of awareness measure (Wiens & Öhman, 2002). However, it is unclear whether the awareness measure used by Morris et al. (1998) assessed the subjective or objective threshold. In the study, participants responded that they saw all of the nonmasked angry faces and none of the masked angry faces. However, because participants were allowed to respond only yes or no to indicate their awareness, they may have reserved the yes response only for the nonmasked angry faces even though the masked angry faces might have been partly detectable. That is, although participants might have been able to discriminate the masked angry faces, they responded no (i.e., unawareness) to masked angry faces because these fell below their subjective criterion of visibility. Thus the 0% correct on the (objective) detection task might be an index of the subjective threshold rather than the objective threshold. This issue is further complicated by the fact that the detection task might index a task dimension that is independent from the dimension of interest (Duncan, 1985). Whereas the detection task assessed participants' ability to detect either angry face, the dimension of interest referred to the difference between the angry face paired with noise versus the angry face not paired with noise. So, even if participants responded that they did not detect either masked angry face, they might have been able to discriminate between the two masked angry faces. Consistent with this idea, Haase et al. (1999) showed that even though participants responded that they did not detect either masked stimulus (CCC or ZZZ), they were able to discriminate between them. Therefore, although the findings of qualitative differences in the Morris et al. (1998) study support the validity of the awareness measure, further research is needed to clarify questions regarding participants' awareness.

When Are Differences Qualitative?

Although findings of qualitative differences for masked and nonmasked emotional pictures provide strong evidence for unconscious processes, it seems difficult to identify emotional processes that are predicted to differ qualitatively for masked and nonmasked pictures. For example, findings that effects are significantly stronger for nonmasked than masked pictures are insufficient for qualitative differences, as effects for masked and nonmasked pictures would merely differ in size, not qualitatively. Although effects that differ in direction are the strongest evidence for qualitative differences, there are scenarios that are less obvious. For example, the findings by Morris et al. (1998, 1999) regarding differential activations in the left and right amygdala suggest that effects were qualitatively different for masked and nonmasked pictures. Similarly, counterintuitive findings of significantly stronger effects to masked pictures than to nonmasked pictures might also represent a qualitative difference (e.g., Rotteveel, de Groot, Geutskens, & Phaf, 2001). However, any claims for qualitative differences need to be carefully examined. For example, Chesman and Merkley (1986) reported that participants were able to use a predictive strategy in a Stroop priming task when words were not masked but were unable to do so when words were masked. The authors concluded that these data provided evidence for qualitative differences. However, because there was no significant effect for masked words and a significant effect only for nonmasked words, these results might represent not qualitative but quantitative differences. Similarly, the findings by Morris et al. (1998) are consistent with the notion of qualitative differences. However, they might also be conceptualized as quantitative differences. For example, the right and left amygdalae might differ in their relative activation functions to degraded and nondegraded input. If so, this phenomenon might be similar to sensitivity differences to frequency for hair cells in the cochlea. Because hair cells vary in their sensitivity to different frequencies depending on their location within the cochlea, cells at one end
of the cochlea are mostly sensitive to low frequencies, whereas cells at the other end of the cochlea are mostly sensitive to high frequencies. Although these sensitivity differences depending on location can be conceptualized as qualitative differences, they might be more parsimoniously described as relative (quantitative) differences. Accordingly, the right amygdala might be relatively more sensitive to degraded input, whereas the left amygdala might be relatively more sensitive to non-degraded input. Even if this alternative explanation might seem unlikely, it illustrates the need to consider whether findings of apparent qualitative differences can be interpreted more parsimoniously in terms of quantitative differences.

The Psychophysical Approach

An alternative to the search for qualitative differences is the psychophysical approach. The main purpose of this approach is to determine the dose-response relationship between awareness and the emotional measure of interest. Thus awareness is not necessarily treated as a dichotomous phenomenon (i.e., participants are either aware or unaware of the masked pictures) but can vary in degree. To manipulate awareness, masked pictures are presented at various masking parameters. Then the relationships between effects of masking parameters on participants' awareness of the masked pictures and the emotional measure can be assessed. Hence, given that issues relating to the concept of thresholds (subjective vs. objective) and the measurement of awareness are far from resolved (as discussed previously), the psychophysical approach allows a more eclectic approach.

The psychophysical approach has several notable features. First, the psychophysical approach provides information about the shape of the response function across a range of masking conditions. Thus, instead of trying to prove that unconscious processes exist, the psychophysical approach provides quantitative information about relative changes in emotional processes as awareness is varied. So, even if the underlying relationship resembles a step function, results from a study covering a range of masking conditions could provide good evidence for this step function. For example, results might show that effects on the variable of interest remain unchanged across the short range of masking intervals, increase in a step-like fashion at the middle intervals, and remain stable for the longer masking intervals. This pattern would support the idea that the underlying relationship follows a step function. In contrast, even if a study with two conditions included the two critical middle intervals, results would not allow one to conclude anything about the effects of short and long intervals, as findings of a step function for the middle intervals would also be consistent with other relationships (e.g., U-shape) across intervals. In fact, as illustrated in Figure 5.9, the relationship between awareness and emotional activation might not be linear, as suggested by some models of unconscious perception (e.g., Snodgrass, Bernat, & Shevrin, 2004). If so, studies that sample only two points (unconscious vs. conscious) will not detect these relationships. Hence, they do not allow one to draw strong conclusions about the underlying relationship between awareness and emotional activations. In contrast, the psychophysical approach can capture the exact nature of the relationship between awareness and emotional activation.

Also, because the psychophysical approach aims at studying the shape of the relationship across a range of conditions, it advocates the use of complementary measures of awareness. As described earlier, there is an ongoing debate on how to measure awareness. Even if awareness is simply classified in terms of subjective and objective thresholds, it is unclear which measures are best at assessing these thresholds. However, because research suggests that the thresholds differ, research that studies effects below and above these thresholds could provide evidence to support either or both concepts. For example, because research suggests that stronger masking (e.g., shorter SOA) is necessary for participants to be unaware at the objective rather than the subjective threshold, a study that covered both thresholds could be useful in determining whether changes in the variable of interest occur when participants become aware at the objective or subjective threshold or both. To illustrate, Figure 5.10 shows hypothetical emotional activation at different levels of unawareness. As shown, in a strong model of unconscious processing, emotional responses occur already below the objective threshold. Alternatively, in a weak model of unconscious processing, emotional responses occur above the objective but below the subjective threshold (for an example regarding amygdala activation to masked fearful faces, see Pessoa, 2005).

Last, because the psychophysical approach covers a range of masking parameters, it minimizes the risk of missing the critical range of masking parameters. Because unawareness effects might occur only within a small range of masking parameters, an experiment with only a single unaware con-

Figure 5.9. Hypothetical dose-response functions between awareness and emotional activation. Because emotional responding might not be linearly related to awareness, manipulation of target duration in small steps permits the study of the nature of the relationship (i.e., dose-response function) between awareness and emotional activation.
awareness of the masked targets (hits) can be contrasted with trials in which participants reported no awareness of the masked targets (misses) to study effects of awareness per se. Of course, this assumes that mere reporting of awareness (false alarms) has no effects (e.g., for amygdala activations for false alarms to fearful faces, see Pessoa, Japee, Sturman, & Ungerleider, 2006). However, although this design can help identify processes that are correlated with awareness, it is also insufficient to prove the causal nature of the relationship between awareness and obtained findings, as awareness could be an epiphenomenon (Frith, Perry, & Lumer, 1999). Nonetheless, because masking research has been mostly interested in determining whether emotional processing can occur in the absence of awareness, effects of awareness per se are irrelevant for this question. Thus, to study which degree of emotional processing can occur at different levels of awareness, the psychophysical approach is a valid strategy.

Example Study of the Psychophysical Approach

A current study in our lab using functional magnetic resonance imaging (fMRI) illustrates the psychophysical approach (pilot results were presented by Wiens, Fransson, Ingvart, & Öhman, 2004). The purpose of this study is to further our understanding of the relationship between awareness and amygdala activation to fearful faces (Whalen et al., 1998). In the study, target pictures were fearful, neutral, and scrambled faces. To manipulate awareness, targets were presented for durations of 15, 20, 30, and 60 ms and were followed immediately (masked) by scrambled faces that were shown for 500 ms minus target duration (so as to have the same picture duration in all conditions). As part of the study, participants (N = 15) performed two recognition tasks in the fMRI scanner. One task was a two-interval forced-choice (2IFC) face detection task. On each trial, a masked fearful or neutral face and a masked scrambled face (lure) were shown with a 1.5-s interval between them. Then participants decided whether the first or the second masked picture was a face and whether they actually saw a face or were just guessing. The other task was similar (also 2IFC) except that a masked fearful face and a masked neutral face (lure) were shown on every trial. In this fear-recognition task, participants decided whether the first or second masked face was fearful and reported whether or not they actually saw a fearful expression. The present study used 2IFC tasks instead of simple (yes/no) detection tasks for the following reason. Because participants often report little, if any, awareness of masked pictures, their subjective sense of awareness might undermine their performance on simple (yes/no) detection tasks. For example, participants might lose motivation and respond yes and no randomly. As a result, performance would be underestimated. In contrast, 2IFC tasks tend to avoid this problem because a target is actually presented on every trial (Macmillan & Creelman, 1991). Hence performance estimates can be expected to be more accurate for 2IFC than detection tasks.
Figures 5.11 shows preliminary results for the face-detection task (A and B) and the face-recognition task (C and D) for the four target durations. Panels A and C are based on signal detection analyses of the data. Panel A shows mean percent correctly detected faces for fearful and neutral faces, and panel C shows mean percent correctly recognized face, expressed as adjusted maximum percent correct (PCmax). Panel B shows mean percent of reported faces for fearful and neutral faces, as well as for scrambled faces (i.e., false alarms), and panel D shows mean percent of reported face for correct and incorrect responses. Results in panels A and C are presented in PCmax rather than raw percent correct because the latter measure can be confounded by response biases. Also, PCmax was used instead of the signal detection index $d'$ because readers tend to be more familiar with percent correct. Accordingly, the signal detection index $d'$ was computed first. However, because sensitivity ($d'$) cannot be computed for null or perfect performance, half a response was either added or subtracted, respectively. Because in these 2IFC tasks, participants could decide based on two intervals, sensitivity estimates ($d'$) were adjusted (divided by $\sqrt{2}$). Then, for each participant, PCmax was defined as percent correct for a given

**Figure 5.11.** Performance on the two-interval forced choice (2IFC) tasks across four target durations ($N = 15$). Panel A shows mean (with 95% confidence intervals) percent correctly detected faces for fearful and neutral faces in the face-detection task, as expressed in maximum percent correct (PCmax) from signal detection (SDT) analyses. Panel B shows the mean (with 95% confidence intervals) percent reported faces for fearful, neutral, and scrambled faces in the 2IFC face-detection task. Panel C shows mean percent of fear that was correctly recognized in the 2IFC face-recognition task (as adjusted PCmax). Panel D shows mean percent reported fear for fearful and neutral faces in the face-recognition task. Note that PCmax is the maximum percent correct without any response biases and adjusted (divided by $\sqrt{2}$) for the fact that participants could decide based on two intervals.

As shown in Figure 5.11, as target duration increased, percent of detected and recognized faces increased gradually rather than in a stepwise fashion. This was true also for psychophysical curves of individual participants (not shown). However, as shown in Figures 5.11A and B, there was no difference in face detection for fearful and neutral faces. Also, across target durations, participants performed better on the face-detection task (A) than on the face-recognition task (C). These findings indicate that face detection was easier than fear recognition. Also, findings suggest that manipulation of target duration was successful in affecting participants' reported awareness and recognition performance of masked faces.

However, when discussing the data across participants in terms of objective and subjective thresholds, the following conclusions might be drawn. Because participants did not perform beyond chance in discriminating between fearful and neutral faces at target durations of 15 and 20 ms (Figure 5.11C) these findings suggest that fear recognition was below the objective threshold at these intervals. However, because face detection was much greater than chance even at 15 ms (Figure 5.11A), the objective threshold for face detection was exceeded in all target durations. In contrast to conclusions about objective thresholds, inferences about subjective thresholds are harder to draw. If the subjective threshold were defined as the target duration at which the percent of reported faces differs significantly from zero (or from the false alarm rate), the subjective threshold would not differ from the objective threshold (because both would be based on discrimination performance). Therefore, a definition of the subjective threshold cannot rely on discrimination performance. One possibility is the use of an empirical threshold (e.g., 50%). If so, participants would be considered unaware below the subjective threshold (50%) at target durations shorter than 60 ms, as they reported less than 40% of the faces in the face-detection task and less than 20% of the fearful expressions in the fear-recognition task (Figure 5.11B and D).

Because fMRI was also measured in a separate part of the study, we can study brain activations to masked faces at different levels of awareness. Results from this study will address whether amygdala activation occurs to masked fearful faces below both objective and subjective thresholds or whether it occurs only above the objective but below subjective threshold (see Figure 5.10). Also, because, contrary to previous research, scrambled faces rather than actual faces served as masks, we can study activations to masked faces per se. Further, although it was not feasible to include more than four masking conditions, we can study dose-response relationships between awareness and emotional processing (see Figure 5.9). That is, instead of dichotomizing awareness in terms of subjective and objective thresholds, performance measures can be used to index awareness on a continuum. Thus the debate about concepts of thresholds can be avoided.

Conclusions

This chapter reviewed conceptual, methodological, and technical challenges in the use of masking to study unconscious emotional processes. Traditionally, the approach has been to try to eliminate awareness to assess effects of masked target pictures of which participants are not consciously aware (i.e., dissociation paradigm). Unfortunately, there is no definition and measure of awareness that researchers agree on (i.e., issues of exhaustiveness and exclusiveness). Also, because unawareness has been indexed as null sensitivity of the awareness measure, it has been defined in terms of a statistical criterion. Thus it depends on statistical power and is also affected by difficulties in proving the null hypothesis (i.e., participants are unaware). Further, because masking requires that pictures be shown briefly and in short order, it places great demands on display technology. Unfortunately, our research suggests that commonly used displays based on TFT and LCD technology do not recommend themselves for masking. Therefore, although some displays can handle masking, each setup needs to be validated. Hence we urge researchers to provide evidence for the validity of their display equipment in their publications. Also, because there is no gold standard for masking parameters and because perceptual processes are affected by small changes in masking parameters, researchers need to focus on assessing awareness in their participants rather than on reporting only the masking parameters.

Because of its focus on conceptual, methodological, and technical aspects involved in masking, this chapter has discussed findings from masking studies on emotional activations only if relevant. For example, there are a number of recent imaging studies on awareness and amygdala activation (for review, see Wiens, 2006). Although we are enthusiastic about this surge in research, we would like to encourage researchers to employ the psychophysical approach described here in their studies. Also, because all of these studies used displays (TFT, LCD) that might have questionable validity for masking, we invite researchers to use a setup similar to ours involving shutters (see Figure 5.6). Because shutters are not limited by refresh cycles, they permit manipulation of picture duration in milliseconds. Thus they provide an excellent means to study the relationship between awareness and emotional activation.

One reason that researchers disagree on how to define and measure awareness might be that it is an oversimplification to treat awareness as a unitary concept. Awareness might have to be conceptualized in terms of processes such as detecting, discriminating, identifying, and noticing. If so, comprehensive measures (e.g., subjective and objective) should be used to characterize these processes. Although findings of qualitative differences represent strong evidence for the distinction between conscious and unconscious processes and for particular measures and concepts of awareness, the psychophysical approach provides an alternative strategy for studying the dose-response relationships between awareness
and emotional processes. When used with masking, these approaches allow researchers to study the role of awareness in emotional processes and thus to probe unconscious emotional processes.

Acknowledgments

We thank Ed Kallin for helpful comments. Stefan Wiens would like to thank his family in Sodericka for a relaxing writing environment during the summer of 2003. The authors’ research was supported by grants from the Bank of Sweden Tercentennial Foundation and the Swedish Council for Research in the Humanities and Social Sciences to Arne Ohman.

Notes

1. Unconscious simply means not conscious.
2. Examples are Presentation (Neurobehavioral Systems), E-prime, ERTS (Beringer, Frankfurt am Main, Germany), and DMDX (http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm).
3. Reingold and Merikle (1988) proposed an approach in which direct and indirect measures are compared in their relative sensitivity. If participants perform better on the indirect measure (e.g., of emotional processes) than on the direct measure (of awareness), then this would be evidence for unconscious processing. Although this is an interesting design, it requires that both measures be assessed under comparable conditions, including the same response metric. Unfortunately, this requirement is unlikely to be met by most measures of emotion.
4. As Cohen (1994, p. 1001) put it: “Go build a quantitative science with p values!”

References


Roser, M., & Gazzaniga, M. S. (2004). Automatic brains,