Polygraph

VOLUME 32 2003 NUMBER 3

Contents

Published Quarterly

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph

Subscription Information: *Polygraph* is published quarterly by the American Polygraph Association. Advertising and Editorial Address is P.O. Box 10342, Ft. Jackson, SC 29207 (USA). Subscription Rates: One year $80.00 (domestic), $100.00 (foreign). Change of address: APA National Office, P.O. Box 8037, Chattanooga, TN 37414-0037.

The Sheridan Press

The Neuropsychology of Emotion Processing and Its Relevance to the Detection of Attempted Deception

Joanna Aycock[1] and Lee H. Wurm[2]

Abstract

In this paper we review the literature on the activation of brain structures and emotion processing from a neuropsychological perspective, attempting to integrate recent findings. Emotion may provide a useful link to the recent work of researchers examining deception using neuroimagery techniques. Definite, qualitatively different patterns of brain activity for the different emotions would prove immensely useful to theorists and applied specialists interested in the detection of deception. It seems unlikely that such patterns will be forthcoming given current techniques and the present state of scientific knowledge. However, it is likely that the imperfect techniques and findings described in this review, together with existing (imperfect) techniques and knowledge, can improve the field substantially.

The human brain is incredibly complex, and although tremendous progress has been made there is significant disagreement about the relationships between the brain, emotions, and deception. With the increasing availability of sophisticated techniques for examining brain activity, recent research is providing valuable clues. However, it is also leading to additional questions.

Deception and its detection involve brain activity, emotion, and sympathetic autonomic arousal, and we believe that people interested in the detection of deception can benefit from what is currently known about brain activity and emotion. Ekman (2001b) lists three emotions as important and intertwined when it comes to deceit: fear about being caught, guilt about lying, and delight in having gotten away with the lie (see also Buller & Burgoon, 1998). However, identification of these emotions can be extremely tricky for several reasons. For example, according to Ekman, the innocent person's fear of being disbelieved is indistinguishable from the guilty person's fear of being caught. In addition, so-called natural liars know that they are good at deceiving because they have been getting away with things for years. Consequently they probably will not have much fear about being caught, and they may not feel much guilt about lying. For reasons such as these, even if it proves possible to link specific emotions to deception, individual (or even subgroup) differences will complicate the picture.

The exact relationship of emotion and deception is even more complex than has been just stated. Emotion can lead to the initial desire to deceive; emotion is frequently a concommitant of deception; and emotion often follows as a consequence of deception (Buller & Burgoon, 1998). This, too, serves to increase the difficulty of relating emotion to deception in a way that can be of practical use to those interested in lie detection. However, the hope is that knowledge of brain activity during emotion processing or during attempted deception can ultimately be incorporated into a "toolbox" used by practitioners, that includes not only standard techniques but also things like voice analysis (e.g.). person's voice. In addition, Ekman (2001a) has found that for 70% of the people tested, vocal pitch is higher under upsetting conditions (Ekman, though, is careful to say that higher pitch is not a sign of deceit; rather, it's a sign of emotion, which may be correlated with deceit).

[1] Joanna Aycock, Department of Psychology and Program in Linguistics. Correspondence may be sent to Joanna Aycock, Department of Psychology, Wayne State University, 71 West Warren Avenue, Detroit, MI 48202. Electronic mail may be sent to [jaycock@wayne.edu].

[2] Lee H. Wurm, Department of Psychology and Program in Linguistics.

Pollina, Vakoch, and Wurm (1998) found that the emotional arousal associated with attempted deception produced small but detectable changes in the acoustics of a Vrij (2000), too, concludes that higher pitch is a relatively reliable cue to deception (because of the emotional stress of lying), and reports that liars have longer pauses in their speech because they are thinking harder.

Another promising avenue of research uses neuroimaging techniques. Some recent studies suggest that activity of specific brain areas can be linked to certain kinds of deceptive behaviors. For example, Langleben, Schroeder, Maldjian, Gur, McDonald, Ragland, O'Brien, and Childress (2002) used the blood oxygen level-dependent technique paired with fMRI scans to record changes in brain activity while their subjects performed a Guilty Knowledge Test. The test required motor responses to simple "yes" or "no" questions. The researchers wanted to find a reliable deception detection technique based on brain activity and uncontaminated (to the extent possible) by the unreliable emotions of the individual subjects. Their subjects, when asked to lie, experienced a significant increase in activity in the anterior cingulate cortex, an area linked in previous studies to the presence and degree of response conflict. Other structures, some of which are involved in initiating and executing voluntary motor functions, also showed increased activity. This indicates an increase in the effort needed to suppress the truth to allow the deceitful response to emerge. Spence, Farrow, Herford, Wilkinson, Zheng, and Woodruff (2001) conducted a study with a similar methodology, and found quite similar results. They reported additionally that subjects took significantly longer to respond in the lying task than in the truth task. This, too, can be explained in terms of response inhibition, and provides another way to link the deception process to something measurable independent of emotion.

Brain imaging techniques, then, may be used to determine which questions result in increased activity in "truth suppression" areas. This could allow practitioners to know with better reliability which responses involve deception. Results such as these are quite promising. As we will see below, some of the same structures identified by Langleben et al. (2002) and Spence et al. (2001) are implicated in neuropsychological studies of emotion processing. This opens up exciting potential avenues of research and allows for provocative links to be established across research areas.

In this paper we attempt to integrate recent neuropsychological findings in the area of emotion processing. We will outline the development of theoretical thinking about the roles of the limbic system and, more specifically, the amygdala, in relation to emotion recognition. First, the different theories concerning emotion and neural architecture will be introduced, including a discussion of the contributions of the right and left hemispheres, and the amygdala and its surrounding structures in emotion recognition. We will explore the possibility that larger integrated neural systems, rather than single structures, participate in emotion recognition. Next, we will discuss the amygdala's other possible functions, and some possible effects of the mode of stimulus presentation. Finally, we will discuss some general methodological problems of research in this area.

Theories of Emotion Processing

There are three common theories of emotion processing in the brain. Right hemisphere theory (e.g., Jackson, 1874, 1880; Mills, 1912a, 1912b) states that all emotion is processed in the right hemisphere whether it be in the cortical, subcortical (Papez, 1937) or limbic regions. Valence theory (e.g., Ley & Bryden, 1979; Reuter-Lorenz & Davidson, 1981; Reuter-Lorenz, Givis, & Moscovitch, 1983) claims that negative emotions are processed in the right hemisphere and positive emotions are processed in the left hemisphere (see Mandal, Tandon, & Asthana [1991] for a variant of this view). The third theory, which we will call neural systems theory (Adolphs, Tranel, Damasio, & Damasio, 1994, 1995), states that each emotion has its own localized area of processing that may or may not extend into both hemispheres and through cortical, subcortical and limbic tissue. We will now examine each of these theories in more detail.

Right Hemisphere Theory

Right hemisphere contributions to emotion processing were discussed by Jackson

(1874, 1880) and Mills (1912a, 1912b). Jackson observed that damage to the left hemisphere did not result in the impairment of the production of emotional words, and Mills observed that right hemisphere lesions did adversely affect emotional expressions. Since then, the focus of investigation has narrowed and the limbic regions of the brain have gained attention. For example, Papez described an emotion-sensitive region in the limbic system in 1937 (the limbic system is a primitive set of brain structures fundamentally involved in learning, memory and emotion). Research of this kind led theorists to focus specifically on the amygdala, a structure within the limbic system, as centrally important in emotion processing.

Borod, Cicero, Obler, Welkowitz, Erhan, Santschi, Grunwald, Agosti, and Whalen (1998) found support for the right hemisphere theory in a study in which they tested a group of right brain damaged subjects, a group of left brain damaged subjects, and a group of normal controls. Borod et al. (1998) presented subjects with eight different states of emotional expression, three positive and five negative. They found that when right brain damaged subjects had to discriminate and identify emotion from facial expressions, prosody (i.e., tone of voice, or emotion expressed vocally), words and sentences, they performed significantly worse than left brain damaged and normal control subjects. These results were found no matter which of the eight emotions was presented.

Further support for the right hemisphere theory comes from Adolphs, Damasio, Tranel, and Damasio (1998a). This research, like that of Borod et al. (1998), involved testing right brain damaged and left brain damaged subjects on a series of facial emotion rating tasks. The results indicated that only damage to the right hemisphere caused emotion recognition deficits. It was also discovered that there were specific brain regions, or hotspots, on the lateral and mesial surfaces the right hemisphere that contributed to emotion recognition. Damage to the right inferior parietal cortex and damage to the anterior infracalcarine cortex caused the most deficits.

Adolphs et al. (1998a) also attempted to isolate those subjects who did not have any subcortical or limbic damage. This is important because conclusions about hemisphere function can be severely compromised if the nature of the brain damage is not specific to the structures of interest. Subjects with primary lesions in the subcortical areas or with lesions in the prefrontal cortex were excluded from the Adolphs et al. (1998a) study, because these areas are known to be closely related to the amygdala and to contribute to emotion recognition.

Lorch, Borod, and Koff (1998) did research on right brain damaged patients, left brain damaged patients with aphasia, and normal controls. The researchers found that left brain damaged patients could produce both posed and spontaneous facial expressions as well as control subjects could. The right brain damaged individuals performed at a significantly lower level than the controls.

Anderson, Spencer, Fulbright, and Phelps (2000) and Adolphs, Tranel, and Damasio (2001) also tested the ability of subjects with either right brain damage or left brain damage on facial emotion evaluation. All subjects had a unilateral anteromedial temporal lobectomy. The damaged right hemisphere was found, in both studies, to lead to deficiencies in emotion recognition. However, both cortical and subcortical regions were damaged in the lobectomy, so, these studies still do not help to discriminate between cortical and subcortical contributions to emotion processing and recognition. The authors concluded that the right hemisphere recognizes general emotions of avoidance-withdrawal, but specific regions within the right hemisphere recognize specific emotions. There was a deficiency in fear evaluation in those subjects with right hemisphere anteromedial temporal lobectomy which includes the right amygdala. Morris, Friston, and Dolan (1997) also found evidence for the selective right hemisphere activation to salient aversively conditioned stimuli. In this study there were many structures activated by the stimuli, but they all were in the right hemisphere.

If all emotion is processed in the right hemisphere and only the right hemisphere, it is natural to conclude that there are not any primary links to emotion recognition in the left hemisphere. Therefore, damage to the left hemisphere should not result in any emotion deficits. As noted, Lorch et al. (1998) found that left brain damaged patients produced emotional expressions comparable to normal controls. However, they also explored the possibility that the left hemisphere is a secondary emotional system capable of compensating for deficits in the primary emotional system found in the right hemisphere.

There is empirical evidence to support the hypothesis that the left hemisphere plays a role in emotion recognition. For example, Stone, Nisenson, Eliassen, and Gazzaniga (1996) performed some interesting work on a split-brain patient. The subject underwent brain surgery to sever his corpus callosum, thus breaking the communication between the two hemispheres of his brain. In a split-brain patient, the primary connection between the cortical hemispheres is no longer functional, and therefore, it is argued that stimulus information presented to one hemisphere cannot travel to the other. The subject was required to label the facial expressions presented separately to both his left and right hemispheres. Stone et al. argued that if the subject was capable of making accurate judgments about the stimulus information presented to his left hemisphere, it would speak to the existence of some emotion recognition capacity in that hemisphere, independent of the right "emotion" hemisphere. The subject was in fact able to accurately label facial emotions at a similar level regardless of hemisphere; therefore, there must be some neural ability in the left hemisphere for emotion recognition. Although the researchers assert that both brain hemispheres have some ability to recognize emotion, they do not claim that both hemispheres use the same processes to recognize emotion.

While the study of Stone et al. (1996) was valuable, some potentially important issues remain unresolved. Perhaps most importantly, the study did not take into consideration the fact that there are other connecting fibers between the left and right hemispheres (and so the widely-used term "split-brain" is really a bit misleading). The anterior commissure, for instance, is an additional group of connecting fibers linking the two temporal lobes (Heimer, 1995) that remained intact in this subject's brain. The amygdala is found within the temporal lobes, which is significant because as we discussed above, the amygdala is commonly believed to be a major emotion-processing center. Many subcortical efferent connections between the amygdala and the other "emotion specific" subcortical organs also remained intact in the subject, with the possible exception of those connected by the hippocampal commissure. Without knowing just how extensive the damage was to the subject's brain, it cannot be said with certainty then that the left cortical hemisphere can recognize emotion.

Nevertheless, the Stone et al. (1996) work has opened the door for other researchers to go a step further and investigate just *how* the left hemisphere may be involved. For example, there has emerged a theory that associates the lateralization of brain function with the valence of the presented emotion.

Valence Theory

A competing theory of emotion recognition, which is increasingly attracting the attention of researchers, is the valence theory. According to this theory, there is a more specialized type of anatomical lateralization in emotion recognition. Instead of all emotion being processed and recognized using the right hemisphere, it has been suggested that the right hemisphere is responsible for the recognition of negative emotions, those associated with avoidance and withdrawal. Positive emotions (and possibly those associated with attack behaviors) are processed by the left hemisphere, according to this theory.

There is some recent evidence in favor of this theory. For example, Adolphs et al. (1998a) tested subjects with right hemisphere cortical damage on facial recognition tasks. They found that these subjects were only deficient in recognizing negative facial emotions. The recognition of happiness was never impaired. This agrees with the

predictions of valence theory. It should also be noted that the damage to these subjects' right hemispheres was limited to the higher sensory cortices and did not involve subcortical structures. Because of this, the authors concluded that the neocortex is necessary for processing negative facial emotions, even if the subcortical structures and prefrontal cortex are intact.

There is other evidence to suggest that valence does influence emotion recognition, but it is not exactly clear how. For example, Anderson et al. (2000) conducted a study with subjects who all had unilateral anteromedial temporal lobe damage and they found that the right brain damaged group was deficient in recognizing happy faces. According to valence theory, recognition of positive emotions should have been impaired in left brain damaged subjects, but not in right brain damaged subjects (but see Mandal et al., 1991). The authors believe that damage to the substantia innominata, and not to the amygdala, was responsible for this pattern of results (this is an important distinction, to which we will return below). There is in fact debate about whether the recognition of happy faces can be impaired. Many studies have suggested that it can (Adolphs et al., 2001; Anderson & Phelps, 2000; Morris, Friston, Büchel, Frith, Young, Calder, & Dolan, 1998b), but others have led to the conclusion that it cannot (Adolphs, Schul, & Tranel, 1998b; Adolphs et al., 1998a; Adolphs, Tranel, Hamann, Young, Calder, Phelps, Anderson, Lee, & Damasio, 1999a; Calder, Young, Rowland, Perrett, Hodges, & Etcoff, 1996).

Hamann, Ely, Hoffman, and Kilts (2002) did research on the pattern of brain activation in response to negative and positive emotional stimuli and neutral stimuli of either low or high interest. Using PET scanning during the presentation of the visual stimuli, the increased brain activation was measured. While there were some minor differences in the specificity of the brain regions responding to the two emotion conditions (positive and negative), the authors did not find any evidence of the activation lateralization that would be expected if the valence theory were true. Positive emotions caused activation in the left amygdala, ventral striatum (a part of the basal ganglia), the ventromedial prefrontal cortex and the visual cortex. Negative emotions led to activation in the amygdala bilaterally but with stronger activation found in the left side. High interest, neutral stimuli, too, caused activation in the left amygdala and visual cortex.

Unlike the majority of research summarized in this section, Hamann et al. (2002) did not use faces as stimuli. This potentially allows for conclusions to be extended beyond the realm of facially-presented emotion, but because the stimuli varied in interest level, it cannot be known for sure what the activated brain regions were reacting to: differences in valence or differences in interest level. In addition, the stimuli were not rated to determine which emotion they expressed and so any brain activation in response to them can only be considered in general terms. Therefore, the brain regions activated can neither be linked absolutely to differences in the affective value of the stimuli nor to any specific emotion presented.

The valence studies have suggested that positive and negative emotion play a large role in determining the pattern of brain activation. However, additional studies indicate that there are other factors that need to be considered (e.g., interest level). Neural systems theory is one attempt to address these additional factors.

Neural Systems Theory

The first adaptation made to the valence theory was the addition of the idea that there are different neural systems that transcend the common hemispheric distinctions associated with the two general classes of affective stimuli. There is an aversive neural system (one that responds to those emotions signifying potentially harmful situations) and there is also an appetitive neural system (one that responds to those emotions signifying a safe and inviting situation).

Lane, Reiman, Bradley, Lang, Ahern, Davidson, and Schwartz (1997b) found that both pleasant and unpleasant emotional facial stimuli led to activation in the thalamus (a collection of nuclei which have connections with the cerebral cortex and other subcortical

structures), the hypothalamus, the medial prefrontal cortex, and the midbrain, as measured by PET scans. Unpleasant stimuli led to additional activation in the left amygdala, the hippocampus, the left parahippocampal gyrus, both sides of the cerebellum, and both sides of the occipito-temporal cortex (some of these structures will be described below). Pleasant emotional stimuli elicited additional activation in the head of the left caudate nucleus. There was activation found on both sides of the brain for the unpleasant stimuli which provides additional support for the idea that valence has some role to play in this theory of emotion recognition as well. There is merit then in not disregarding valence information altogether. This new theory, however, is different in that it does not claim that the system of activated structures are all within one hemisphere.

Adolphs et al. (1999a) included a larger subject group than previous studies of this type. Nine subjects were tested, all of whom had been previously used in other studies. This larger subject group allowed the researchers two advantages: comparison of the methodologies of previous studies on facial emotion recognition and an increase in the statistical power of the results, thereby allowing for stronger comparisons and greater confidence in the conclusions. Adolphs et al. found that damage to the amygdala did not inhibit fear recognition specifically but the recognition of a more general class of emotions signifying threat or danger. They did find large inconsistencies in the nine subjects' ability to rate fearful and angry faces. Despite these inconsistencies, however, the authors concluded that there are systems of neural structures, not just individual organs, underlying emotion recognition.

In other studies, however, valence has not been factored into discussions of neural systems. For example, Adolphs et al. (1998a) found that only structures in the right hemisphere participated in the recognition of facial emotion. As noted above, the researchers excluded participants with subcortical and/or prefrontal damage, and so all of the involved structures were in the higher sensory cortices. Not only was there evidence for primary right hemisphere involvement in emotion recognition but also the participation of specific brain regions within the right hemisphere (e.g., right inferior parietal cortex; anterior infracalcarine cortex). Like the Hamann et al. (2002) study, this suggests that concepts such as "right hemisphere" and "valence" are too broad to provide satisfactory explanations of emotion recognition.

Is the Amygdala a Fear-Specific Organ?

For many years researchers have had a very narrow view of which organs played a role in emotion recognition. The amygdala has long been suspected as the primary executor of fear recognition and since very early in this line of research on emotion processing and recognition, this brain region has been the focus. Recent research has demonstrated, however, that this does not give a complete picture of what is happening.

For example, Morris et al. (1998b) found that structures in addition to the amygdala were active in response to fear. Normal subjects had to rate facial expressions of increasing intensity, as either happy or fearful. Using PET scans, Morris et al. found that the left amygdala did respond significantly to an increase in fearful stimuli, however, significant activation was also found in the left pulvinar (one of nuclei found within the thalamus, which is active during visual perception and language processing), the right anterior cingulate and left anterior insula as well.

Adolphs et al. (1999a) tested nine brain damaged subjects on facial emotion recognition and concluded that the amygdala is not a fear-specific organ but that it is one structure in a system that plays a role in emotion recognition of facial expressions. Adolphs et al. (1999a) did find that the amygdala participates in fear recognition; however, the amygdala also participates in the recognition of other emotions. This is consistent with the conclusion of Calder et al. (1996), who found that some subjects with amygdala damage not only had problems recognizing fear but, also sometimes anger and surprise.

The architecture of the amygdala is complex, though, and it may be more useful to treat it as a heterogeneous collection of related

structures instead of as a single homogeneous entity. It has many distinct neural cell groupings (Heimer, 1995), and it could be that a particular subdivision of the amygdala is responsible specifically for fear recognition. This would explain why the Morris et al. (1998b) and Adolphs et al. (1999a) contradict the commonly held ideas about the amygdala and fear recognition. When studies are performed on brain damaged subjects to ascertain the extent of amygdala participation in emotion recognition, careful attention must be given to the degree and location of amygdala damage. In addition to the separate amygdala subparts, attention must be given to whether the damage is unilateral or bilateral. Unilateral damage has been found in some research not to result in any deficits in emotion recognition; for example, Adolphs et al. (1995) found that only subjects with bilateral damage to their amygdala performed significantly different than normal controls.

However, other research suggests that unilateral damage may lead to impairment, if the damage is located in a crucial place, and if the damage is extensive enough. Adolphs and Tranel (1999) did find that severe unilateral damage to the right amygdala can cause emotion recognition impairments. Brain-damaged subjects who participated in the Adolphs and Tranel (1999) study had either unilateral left, unilateral right, or bilateral amygdala damage. There was also a group of control subjects who had brain damage not involving the amygdala. Participants were tested on their ability to recognize emotion in prosody (tone of voice). The researchers found that subjects with left unilateral damage performed as well as the control subjects. Those with right unilateral damage showed a mixed pattern. One participant, who had 50% damage, performed normally; a second right brain damaged subject, with 100% damage, was significantly impaired. The subjects with bilateral damage also had significant impairment, but because of the patterns of the unilateral subjects, it is possible that the right-side damage in these bilateral subjects is in fact causing the impairments.

Many other researchers have found additional evidence that unilateral damage to the amygdala affects emotion recognition. Anderson et al. (2000) and Adolphs et al. (2001) found that in facial emotion recognition tasks, subjects with right temporal lobectomies had impairments but those with left temporal lobectomies did not. As we have previously cautioned, though, temporal lobectomies affect many different structures which have all been linked to emotion processing and recognition. More specifically, not only cortical regions but subcortical structures like the hippocampus (most notably responsible for memory functions), parahippocampus, and amygdala are affected by the temporal lobectomies. Although it was claimed by the researchers that both higher and lower functioning structures work together in some fashion for emotion recognition, the multitude of damaged structures that they base this conclusion on also makes it difficult to verify.

Many times, studies that use neuroimaging techniques like PET scanning show an asymmetrical activation pattern in the amygdalae with a disproportionate amount of activation in the left amygdala. Blair, Morris, Frith, Perrett, and Dolan (1999) found that sad facial emotions elicited activation in the left amygdala but that angry faces did not. Similarly, Morris et al. (1997), Morris, Öhman, and Dolan (1998a), and Morris et al. (1999) all found that the amygdala's response to conditioned stimuli was asymmetrical.

In previous sections evidence for a distinction between the right and left hemisphere was introduced. Here, we have presented evidence for a distinction between the left and right amygdalae. There are other amygdala subdivisions, however, in addition to simple left and right distinctions. Whalen, Rauch, Etcoff, McInerney, Lee, and Jenike (1998), for example, found evidence that the dorsal and ventral partitions of the amygdala respond differently to masked fearful and happy stimuli. The ventral portion showed the typical decrease in activation to the happy stimuli and an increase in activation to the fearful stimuli. The dorsal portion showed an increase in activation in response to both happy and fear stimuli, although, its activation increased more in response to the fearful emotional stimuli. Future research in this area should be directed at uncovering the specific roles of these smaller subdivisions.

Other Structures and Other Emotions

Not only do the higher order sensory regions and the amygdala contribute to emotion recognition, but other subcortical and limbic structures do as well. These include most notably the hippocampus, parahippocampus, cingulate gyrus, thalamus, insula, anterior commissure and substantia innominata.

Morris, Frith, Perrett, Rowland, Young, Calder, and Dolan (1996) found that the right medial temporal gyrus, right putamen, left superior parietal lobule and left calcarine sulcus all were significantly activated in response to happy facial expressions when contrasted with fearful facial expressions. In a subsequent study, Morris et al. (1998a) found that activation of specific brain structures was related to increases in intensity for happy facial expressions. The posterior, bilateral striate cortex, the bilateral medial occipitotemporal gyri (lingual gyri), the bilateral lateral occipitotemporal gyri (fusiform gyri) and the superior temporal gyrus all showed significant increases in activation as a function of increased intensity of happiness. Other, higher cortical structures have been shown to be relevant, as well. As noted above, a study by Hamann et al. (2002) found that positive facial expressions caused activation in the left amygdala, ventral striatum, the ventromedial prefrontal cortex and the visual cortex.

Lane, Reiman, Ahern, Schwartz, and Davidson (1997a) conducted an ambitious study in which they examined the PET scans of normal subjects listening to the scripts of either emotionally laden film clips or to scripts of previously collected autobiographical scenarios. The object of listening to the scripts was to induce a target emotion, either happiness, disgust, or sadness, and there was also a neutral emotional condition. They found, consistent with previous research on neuroanatomy, that there was not a simple relationship between anatomical structures and the emotions that they process. They found that there were some areas of activation that were common to multiple types of emotion regardless of valence, and that there were other areas of activation that were singular in their affective connection. Specifically, they found that happiness, disgust, and sadness were correlated with significant activation in the medial prefrontal cortex, the thalamus and certain areas in the anterior temporal region. Happiness was also correlated with activation in the middle and posterior temporal cortex and the hypothalamus. Also a collection of nuclei, the hypothalamus is most notably responsible for the regulation of homeostasis and hormone production and release. Disgust was additionally correlated with activation in the midbrain while sadness was also correlated with activation in the middle and posterior temporal cortex, hypothalamus, the cerebellum, the midbrain, the putamen and the caudate nucleus. The putamen and caudate nucleus together comprise the striatum which is a subpart of the basal ganglia (described more fully below). These results provide evidence that emotions are recognized via systems of both general and emotion-specific neural structures that communicate with one another.

Other research is also consistent with this notion. Phillips, Young, Scott, Calder, Andrew, Giampietro, Williams, Bullmore, Brammer, and Gray (1998) found that there were different brain regions participating in the recognition of fearful and disgusted faces, but also that the hippocampi were activated by both. In this study the amygdala did not appear to be the most activated region of the brain. Blair et al. (1999) found that sadness and anger both activated the anterior cingulate cortex and the right temporal pole. Sadness also induced activation in the left amygdala and right inferior and right middle temporal gyri. Anger elicited responses from the right orbitofrontal cortex. As in other studies of this type, there were regions of both common and discrete activation.

These findings are consistent with the results of Anderson et al. (2000). They found that right brain damaged subjects with significant temporal lobe destruction were deficient in recognizing emotion in faces, including happy faces. Anderson et al. (2000) concluded that because the substantial innominata is so close to the amygdala, it is likely that this structure, too, was damaged, and that this may have caused the impairment in happy face recognition. This would be consistent with previous research showing

that the substantial innominata is important for happiness recognition.

What must be investigated now is the relative contribution of the left and right cortical areas vs. the subcortical areas of the brain in emotion recognition. As just noted, emotion processing often involves multiple brain structures, and it is also the case that most brain structures are involved in more than one specific kind of processing. In addition, as has previously been pointed out in connection with the studies by Stone et al. (1996) and Adolphs et. al. (1998a), the extent of participation of each area remains unknown.

Karow, Marquardt, and Marshall (2001) investigated this issue. To uncover the degree to which each of these brain regions participates in emotion recognition, four groups of subjects were included: left brain damaged subjects with subcortical damage and without, and right brain damaged subjects with subcortical damage and without. The study required subjects to interpret and process emotion encoded in either printed material, facial expressions or prosody (tone of voice). The researchers discovered that when damage was restricted to the cortical regions of the brain, there were no impairments in emotion recognition. When subjects had right hemisphere damage accompanied by subcortical damage they were significantly deficient in their ability to recognize emotion in facial expressions and prosody. Left hemisphere damage accompanied by subcortical damage led to deficiencies in emotion recognition in print and prosody. These results indicate that cortical damage alone is not sufficient to cause emotion recognition deficiencies (at least the kinds of deficiencies assessed in this study). It is interesting to note that both left brain damaged and right brain damaged subjects had similar subcortical damage, i.e., basal ganglia, insula and other subcortical structures. Because cortical damage alone was not found to be responsible for impairments in emotion recognition and because the same subcortical structures cannot be responsible for different recognition abilities when linked to different hemispheres, it stands to reason that the connections between the cortical and subcortical regions are the key element in the differing abilities of the left and right hemispheres with accompanying subcortical damage.

The approach used in the Karow et al. (2001) study is a valuable one, but the researchers would have been able to draw stronger conclusions if they had included fear, disgust, and surprise. The recognition of these emotions has been shown to be deficient in certain subjects with subcortical damage and should have been included as stimuli. Subjects in other studies have been found to be deficient in recognizing anger, disgust and fear. Karow et al. also failed to include a subcortical damage group without the additional cortical damage, which would have been useful for comparison.

Other researchers have found that emotion recognition is dependent on both cortical and subcortical areas of the brain (e.g., Anderson et al., 2000). Sprengelmeyer, Rausch, Eysel, and Przuntek (1998), using functional magnetic resonance imaging (fMRI), found that both the left and right hemisphere and subcortical regions were activated during the processing and recognition of some emotions but not others. Borod et al. (1998) also found in their study that there was no effect of lesion level. In other words, the subjects with subcortical damage did not perform significantly different than those subjects without subcortical damage. This leads to the conclusion that both cortical and subcortical areas must be involved in emotion recognition.

What can be concluded from these results is that the pattern of activation elicited from brain structures during emotion recognition depends on which emotion is being recognized. There are many brain structures that participate in emotion recognition, some that may contribute to the recognition of a specific emotion and others that may contribute to the recognition of several emotions. This constitutes a system of structures that operate together for emotion recognition. Because of this, many current researchers now believe that the amygdala is just one organ in a much larger system of emotion recognizing organs and that cortical, subcortical and limbic brain regions all work

together to create a comprehensive emotional experience.

Evidence That Neural Systems Really Are Systems

One early study that includes the idea of neural systems underlying emotion recognition is Adolphs et al. (1994). Their subject had damage primarily confined to the bilateral amygdala, with some additional damage to the left entorhinal cortex (see Hamann, Stefanacci, Squire, Adolphs, Tranel, Damasio, & Damasio, 1996). The entorhinal cortex (Brodmann's area 28) is a part of the parahippocampal gyrus and has direct neural connections with the hippocampus via the subiculum (together, the entorhinal cortex, parahippocampus, hippocampus, subiculum and the dentate gyrus all make up the hippocampal formation). The entorhinal cortex and the subiculum both contain important efferent and afferent connections between the hippocampus and other subcortical and cortical structures in the brain. Their subject was impaired when recognizing prototypical facial expressions of fear, anger, and surprise, but it is not known which structures are actually responsible for this impairment. It could have been a damaged structure or an intact structure dependent on one that was damaged. That is, not only are multiple brain structures implicated in emotion recognition, but in addition, these structures are often very directly physically connected to one another. These tight connections create difficulties when trying to explain exactly which structures are responsible for emotion recognition. This interconnection of structures and their functions is an example of how the brain is organized not in isolated discrete sections, but in interconnected, networking units. It was in fact ideas such as these, supported by subsequent empirical studies, that led to the formulation of the neural systems theory.

Morris et al. (1998b) studied intensity discrimination in fearful and happy facial expressions. They found that both fearful and happy emotional expressions activated certain common regions of the brain, such as the left occipitotemporal sulcus, the left pulvinar and the right orbitofrontal cortex. Additionally, each emotion independently activated other brain regions. All of the activation produced by the fearful and happy stimuli (i.e., the independent as well as the common activation) was bilateral. The fearful stimuli activated a group of organs that have been found linked together in a system by reciprocal connections. The authors suggest, in accordance with previous findings, that the connections between the pulvinar and the amygdala support the idea of a system for mediating salient visual fearful stimuli. The anterior cingulate is responsible for mediating responses to painful stimuli and has reciprocal connections with the amygdala, pulvinar and the anterior insula (which is beneath the cerebral cortex and functions in some autonomic tasks). The anterior insula also is responsible for mediating responses to negative stimuli and has reciprocal connections to the amygdala. These established anatomical connections and proposed associations with the fearful emotional stimuli presented suggest that there are indeed neural systems (even emotion-specific neural systems), and that they transcend hemispheric divisions. The authors suggest that there is in fact a general avoidance-withdrawal system (see also Adolphs et al., 2001; Anderson et al., 2000). However, this suggestion should be viewed with caution because only happy and fear were tested.

Blair et al. (1999) used PET scans to attempt to locate those structures responsible for the processing of anger and sadness. They too found evidence for the presence of neural systems, in contrast to the idea of a one-to-one relationship between a particular emotion and a particular structure. Sad facial expressions activated the right inferior and right middle temporal gyri and the left amygdala and both sad and angry facial expressions activated the anterior cingulate and the right temporal pole.

Sprengelmeyer et al. (1998) found that when subjects were shown fearful facial expressions and their brain activity was monitored with fMRI, the amygdala did not significantly respond. They found that there was significant activation only in the left frontal lobe and the right fusiform gyrus. The researchers also found activation in both cortical and subcortical brain regions for the recognition of certain emotions. For example, disgust activated the right anterior putamen,

right pallidum, left anterior insula and some areas in left inferior frontal cortex. Anger activated the left inferior frontal cortex, the left posterior temporal lobe and the right posterior gyrus cinguli. This, like the other studies summarized in this section, gives support for the idea that there are different neural systems for different emotions.

There was one area activated by all three of the emotions presented by Sprengelmeyer et al. (1998), the left inferior frontal lobe (Brodmann's area 47). The researchers suggest that although there are separate, emotion-specific, neural systems to process different emotions, there are also some areas responsible for the integration of separately-processed information into a unified whole. The left inferior frontal lobe may be one of those areas. They report that it is a common endpoint for the independent neural systems.

There may be a very simple explanation for Sprengelmeyer et al.'s (1998) failure to observe activation in the amygdala in response to the fearful stimuli as was expected. The researchers point to the possibility that the amygdala became habituated to the negative stimuli (see also Whalen et al., 1998). The fact that they only used negative (anger, disgust and fear) and neutral stimuli but not any contrasting positive stimuli could be a reason why the amygdala may have habituated quickly and not shown any significant response to the presented stimuli.

The Amygdala's Other Possible Functions

Subjects who sustain damage at an early age have different patterns of impairments from those who sustain damage when they are older. Early damage can lead to impairments in recognizing displayed emotion but damage later in life does not always lead to such impairments. This has led some theorists to believe that one of the amygdala's primary functions might be to mediate the emotion encoding process during a person's early years (Adolphs, Russell, & Tranel, 1999b; Calder et al., 1996; Hamann et al., 1996). This represents a sharp departure from the long-held view that the amygdala was important only for the recognition of emotions (e.g., Adolphs et al., 1995). Cahill, Haier, Fallon, Alkire, Tang, Keator, Wu, and McGaugh (1996) recorded normal subjects' amygdala activity using PET scans while the subjects viewed an emotional video or a neutral video. Three weeks following this initial procedure and without forewarning, they questioned the subjects on what they remembered about the films. Cahill et al. found that activation in the right amygdala did indeed correlate with the encoding and long-term recall of the emotional videos but not the neutral.

It has also been claimed that the amygdala's function in emotion recognition is to rate the intensity of the emotion experienced. Adolphs et al. (1995) found that their subject with damage to both amygdalae did not have difficulties recognizing the emotion types but had significant difficulties rating the intensity of certain emotions like anger, fear and surprise. Adolphs et al. (1999b) found that the subject gave intensity ratings that were 4-5 standard deviations below what normal subjects gave for anger and fear. The researchers offer the explanation that damage to the amygdala impairs one's ability to learn declarative knowledge about emotions. According to the authors, the intensity or arousal factor is a quality about a displayed emotion that is used to distinguish it from other emotions.

The studies summarized above have focused almost exclusively on the effects of visually-presented stimuli. We now turn our attention to the possible effects of differences in the mode of stimulus presentation. Are the same brain structures involved in emotion recognition when stimuli are presented auditorily? We also summarize evidence on the effects of subliminal visual stimulus presentation.

Mode of Presentation

As discussed above, the recognition of emotion presented in facial expressions can involve systems of discrete anatomical structures working independently and also some cooperating in integration areas. Do emotional stimuli presented vocally function the same way? In the last decade or so, many researchers have investigated this issue. They have discovered that the two methods of presentation are not the same and do, in fact, affect neural processing in distinct parts of the brain.

Phillips et al. (1998) not only discovered that there were different brain regions activated by facial expressions of fear and disgust but that there were also distinct neural structures activated by the vocal expressions of fear and disgust. fMRIs showed increased activation in several anatomical structures. Brodmann's area 32 showed increased activation for both emotions with auditory presentation, and Brodmann's area 37 showed increased activation for both emotions with visual presentation. Brodmann's areas 22 and 42 showed increased activation for both emotions and with both modes of presentation.

Calder, Keane, Manes, Antoun, and Young (2000) looked specifically at disgust and found evidence for a potential insula-striatal system for recognizing disgust in all sensory modalities tested. Their subject had damage to his putamen and insula, and was tested under four different conditions. Specifically, there were two sets of stimuli using emotional facial expressions, one set of auditory (non-verbal) emotional stimuli, and one set of stimuli in which he had to recognize prosodic emotion. The subject was found to be impaired in disgust recognition regardless of the mode of presentation.

Adolphs et al. (2001) found that in addition to the general hemispheric distinctions, already mentioned, there were differences in which brain hemisphere participated in emotion recognition as a function of the mode of stimulus presentation. The right hemisphere played a significant role in the visual processing of facial expressions; right brain damaged subjects performed at a significantly lower level than the normal control subjects. However, there was not a significant difference in the performance of the subject groups in recognizing emotion conveyed prosodically. The researchers point out that both the left brain damaged and right brain damaged subjects and the controls performed similarly; even though fear was an emotion that gave subjects problems during the recognition task, it was a problem universal to all three subject groups.

Some articles introduced in other sections also contain information relevant to this current topic. For instance, Borod et al. (1998) investigated right hemisphere emotion perception presented in multiple modalities. The researchers did find differences in the performance ability of the groups as a function of the presentation mode. With normal control subjects, the researchers found evidence for a single, unified, emotion processor. Regardless of the mode of presentation, normal control subjects showed a pattern of highly correlated results. The left brain damaged subjects had less significantly correlated results and the right brain damaged subjects had results that were not at all significantly correlated between the modes of presentation. According to Borod et al. (1998), these results indicated that the brain damaged subjects did not use a single, central, emotion processor but relied on "redundant, backup" neural structures, that were more mode-specific. There is then evidence for a central processor of emotion which actually is made up of several redundant, networking subsystems that are revealed if there is damage or excessive stress on the tightly knit unit.

Karow et al. (2001) used multiple presentation modes to investigate what contributions each cortical hemisphere and the subcortical brain regions make to emotion recognition. Participants were asked to recognize emotions presented in facial expressions, or in a linguistic task, or in prosody. The researchers discovered that, except for those subjects with cortical damage alone, the emotion recognition deficits depended on the mode of presentation. The subjects with damage solely to the cortical regions of the brain (left or right hemisphere), did not perform significantly worse than the controls. The researchers report that only the left hemisphere damaged subjects with additional subcortical damage performed significantly worse than all other subject groups during the linguistic task; and that only the right hemisphere damaged subjects with additional subcortical damage performed significantly worse than all other groups during the facial expression recognition task. Both the left hemisphere damaged and right hemisphere damaged subjects with the additional subcortical damage also performed significantly worse than the normal controls and the cortical damage only groups on the prosodic task.

There have also been some studies using the auditory mode of presentation only. Ghika-Schmid, Ghika, Vuilleumier, Assal, Vuadens, Scherer, Maeder, Uske, and Bogousslavsky (1997) did research on a male subject with extensive bilateral hippocampal damage and some slight damage to the basal ganglia. This damage resulted in a significantly impaired ability to recognize auditory fear, but did not result previously in an impaired ability to recognize fear in facial expressions. However, it should be noted that the subject did have significant impairments when recognizing contempt and surprise in facial expressions but was not tested on these emotions in the auditory mode. For comparison, it would have been interesting to test the subject's ability to recognize these emotions in the auditory mode.

Scott, Young, Calder, Hellawell, Aggleton, and Johnson (1997) performed a similar study, but using a subject (DR, who has also been studied by other researchers) with damage to the amygdala and some damage outside of the amygdala. In previous studies, this subject showed significant deficiencies in facial fear and anger recognition; however, Scott et al. (1997) wanted to know if she would be deficient in recognizing the same emotions presented in the auditory mode. They found that she did, in fact, have significant difficulties in recognizing anger and fear in the auditory modality. However, because of her damage external to the amygdala and her inability to identify familiar voices or judge between similar voices, the interpretation of these results should be treated cautiously.

Contrary to the results of Scott et al. (1997), Anderson and Phelps (1998) and Adolphs and Tranel (1999) found that damage to the amygdala did not impair the recognition of auditorily-presented fear. Anderson and Phelps (1998) tested an individual with extensive damage to the left and right hemisphere amygdala, hippocampus, parahippocampus and other efferent projection fibers. They found that the subject had no deficiencies in processing and recognizing auditory expressions of fear. Finally, Adolphs and Tranel (1999) tested several groups of subjects, those with bilateral amygdala damage, those with unilateral left or right brain damage, and brain damage controls. They also found that the bilaterally brain damaged subjects had no significant impairments in emotion recognition in prosody. This indicated that the amygdala activation is not essential to process emotion in all modes of presentation. Two of the participants in Adolphs and Tranel (1999) showed a different pattern, which is in fact consistent with that showed by DR (Scott et al., 1997). Adolphs and Tranel (1999) suggest that the performance of all three of these participants was due to some other, confounding, damage external to the amygdala itself, like the basal ganglia.

There is evidence, then, that an additional detail should be factored into the discussion on neural systems. Not only do different valences and the individual emotions themselves factor into this discussion, but now there is evidence that the different modes of presentation must be considered. It would be worthwhile to replicate existing research on visual processing of emotion using other presentation modalities as well. One further example of why this would be useful is the Adolphs et al. (1998a) study. They only looked at emotion recognition in facial expressions and found that damage to the occipital lobe played a role. This is perhaps unsurprising because the occipital lobe is where the primary visual cortex is found. If this study were carried out again using auditory emotional stimuli, damage to the temporal lobe (where the primary auditory cortex is found) might be associated with emotion recognition difficulties.

Conscious vs. Unconscious Perception of Stimuli.

Communication does not have to be mediated explicitly; social signals may be sent and received without conscious knowledge. In an effort to learn how consciousness influences the brain's ability to process and recognize emotions, some researchers have performed studies using masking and aversive conditioning techniques.

Morris et al. (1998a) looked specifically at the amygdala in a study that paired brain imaging techniques with a backward masking procedure. An emotional facial stimulus was subliminally presented (duration = 30 msec)

and immediately replaced by a neutral facial stimulus (i.e., the mask, with duration = 45 msec). In a pre-testing learning phase of the experiment, Morris et al. (1998a) paired an angry face with a blast of white noise. This aversively conditioned stimulus was then presented in either a masked or an unmasked condition, during which time brain images were taken. They found that the amygdala had an asymmetrical activation pattern; the right amygdala was more activated in response to the masked condition but the left amygdala was more activated in response to the unmasked condition. So, there is a difference in brain activity to different levels of emotional awareness. The left hemisphere of the amygdala was more responsive when the subjects were overtly aware of the stimuli's presence, but the right hemisphere amygdala responded more to subliminal stimuli presentations. In fact, according to the researchers, when subjects are able to consciously process the stimuli they can use linguistic processes (which are commonly used by the left hemisphere in general) to boost perception. These lateralization findings were subsequently supported by Morris, Öhman, and Dolan (1999).

Whalen et al. (1998) also tested whether the amygdala can contribute to the recognition of emotion without consciousness. They paired the blood oxygen level-dependent technique with fMRI in either a masked happy or a masked fearful face condition to see which stimulus group would produce the greater level of activation. The target stimulus was presented (subliminally) for 33 msec and the masking stimulus was presented for 167 msec. The general finding was that activation in the amygdala increased significantly to masked fearful faces and decreased significantly to masked happy faces. This gave support for the idea that the amygdala is able to process and recognize emotions even if they are not explicitly apparent. However, there was some variation in the response of the amygdala to the different conditions. The ventral region followed the general activation pattern with consistency. The dorsal region, however, became increasingly activated in response to both fearful and happy faces, although, the activation was much greater in response to the fearful expression.

Activation also extended into the substantia innominata and the anterior commissure which is considered part of the "extended amygdala" (Heimer, 1995). As noted by Whalen et al. (1998), this structure is believed to be activated in response to both fearful and happy stimuli. Therefore, it is possible that activation of the substantia innominata could be mistaken for activation of the dorsal portion of the amygdala in this kind of study. Consideration of the role of the substantia innominata in emotion recognition may also allow us to reconcile the apparently contradictory results about how happiness is processed. As we noted above in the section on valence theory, some theorists have concluded that recognition of happiness can be impaired, while others have concluded that it cannot. However, it could be that happiness activates the substantia innominata, while fear activates the amygdala. Because these structures are so close to each other, it is difficult to be sure that only one or the other has been damaged (and damage to one could be accompanied by damage to the other). It is correspondingly difficult to be sure which of the structures is responding (if indeed only one is responding) to a stimulus.

Whalen et al. (1998) also introduced the idea that it is not conscious awareness of the emotional stimuli that the amygdala is responding to but a more general perceptual salience or arousal value of the stimuli. Some researchers (e.g., Anderson et al., 2000) have used this idea as a possible explanation for their findings, and others have tested perceptual salience directly. For example, through aversive classical conditioning Morris et al. (1997) taught subjects to associate certain facial expressions (presented above the perceptual threshold) with a blast of white noise. Using a PET scan, they found increased activation in several structures including the pulvinar nucleus (found within the thalamus) in response to the conditioned facial expressions. The pulvinar nucleus is closely related to the amygdala and therefore strongly correlated with the amygdala's response to stimuli based on increased salience. Morris et al. (1998b) found that in response to fearful stimuli, both the left amygdala and the pulvinar nucleus were activated. Both are included in a system of other related

structures activated in response to fearful stimuli (see also Morris et al., 1999).

Dimberg, Thunberg, and Elmehed (2000) looked at the unconscious perception and reaction to emotion presented in masked facial expressions. Three sets of masked facial expressions were used: neutral-neutral, happy-neutral and angry-neutral. Masked targets were subliminally presented for 30 msec, followed by the masking stimulus for 5 msec. The emotional response was measured in the activation of the zygomatic major and corrugator supercilii facial muscles. The authors' prediction was that the zygomatic major muscle would increase its activity in response to perceived happy faces. In fact, there was a larger zygomatic major response to the happy masked facial stimuli and a larger corrugator supercilii response to the angry masked stimuli.

Dimberg et al.'s (2000) result provides additional evidence for the unconscious perception, processing and recognition of facial emotion. Together with the other research summarized in this section, this suggests that brain responses to subliminally-presented information (which can be true or false; related or unrelated to criminal activity; etc.) might contain important clues about the truthfulness of the observer.

Expression vs. Reception of Emotion

The research discussed to this point has focused on the reception (i.e., recognition) of emotion. A small literature exists that attempts to document ways in which the *expression* of emotion is similar or different from the reception of emotion. Marquardt, Rios-Brown, Richburg, Seibert, and Cannito (2001), for example, did a study on subjects with traumatic brain injury and their ability to comprehend and express emotion. They found that their traumatic brain injury subjects not only had difficulty identifying emotion but producing it as well (in making facial expressions). However, Marquardt et al. (2001) used an unusual procedure for judging the facial expressions subjects produced. Three people judged each facial expression, but not independently. If at least two of the three judges did not agree they re-viewed the video-taped facial expression and "collaborated to arrive at agreement." Re-viewing and collaborating was not in fact needed for very many of the trials, but the procedure nevertheless raises concerns as to the true extent of the judges' agreement because the potential for influence existed even on trials for which there was apparent agreement. This makes it difficult to evaluate the conclusion that subjects with traumatic brain injury had deficiencies in expressing emotion (see also Spell & Frank, 2000).

Adolphs et al. (1995) tested a subject and discovered that in addition to her documented deficiency in the reception of emotion (i.e., in recognizing emotional input from facial expressions), she could not draw fearful faces. This was, however, not the result of a disabled knowledge of fear; she had a normal concept of fear and was able to describe fear-inducing situations.

Baum, Pell, Leonard, and Gordon (1997) also found that individuals can be impaired in both reception and expression of emotion. They tested the ability of left and right hemisphere damaged subjects to both produce and use prosody to facilitate spoken language comprehension. They found that both the left brain damaged subjects and the right brain damaged subjects were significantly impaired in the reception of prosody and were also both aberrant in their production of prosody.

Anderson and Phelps (2000) performed a study using a subject who had damage to both her left and right amygdalae. She was asked to evaluate emotion in facial expressions, identify emotion in printed form, and generate her own facial expressions depicting target emotions. She had extensive limitations in recognizing emotions in faces or in print (although she did recognize angry and surprised faces within normal ranges). Despite these receptive deficiencies, however, she did not have any deficits when asked to express emotion in facial expressions herself. She did so with normal accuracy on all basic biological emotions tested: fear, anger, surprise, disgust, happiness and sadness. While this finding appears to conflict with that of Adolphs et al. (1995), one must consider that extra damage to the subject's brain in Adolphs et al. (1995) may have had some part to play. In addition, the expressive tasks used were quite different

in the two studies (drawing an emotional facial expression vs. making a face). Because of the small number of studies looking at emotion expression, and the fairly large methodological differences between them, it would be very useful to have additional, well-controlled studies in this area.

Limits of the Current Methodologies

In this section we will discuss the methods used in the research we have summarized, focusing on some of the problems inherent in research of this kind. Many of these problems are due to the fact that large numbers of brain-injured subjects are not readily available (fortunately). Furthermore, each subject comes with his or her own unique pattern of damage, which can severely limit the certainty of the conclusions researchers are able to draw.

One general concern has to do with the possible relationships between the age of a subject at testing, the age of a subject at the time of injury, and the time interval between injury and testing. Some research of this type assumes that brain function has stabilized within several months after the occurrence of a brain injury, but the situation may be much more complex than this suggests. As we noted above, the age of a person when a brain insult occurs may be a major factor, and may have as much to do with subsequent functioning as the damage itself (Anderson et al., 2000; Adolphs et al., 1995, 1998b, 1998a, 1999a, 1999b; Calder et al., 1996; Cahill et al., 1996; Hamann et al., 1996). These factors may also interact with the ages of the research participants at the time of testing. For example, any cognitive/emotional processing declines that accompany normal aging may follow a different course in subjects with brain injuries. Insult to one brain structure is often compensated for by other brain structures, so when normal cognitive declines occur in these compensatory structures, there can be a drastic overall decline in the functioning of a brain-injured subject due to the injury.

A second concern related to the scarcity of brain-injured research participants has to do with limited sample sizes (e.g., Phillips et al., 1998). For example, the Borod et al. (1998) used only 11 right brain damaged subjects and 10 left brain damaged subjects. Compared to some research of this kind, this is a quite respectable sample size. However, each subject had their own unique lesion pattern in the brain, which led to a far greater diversity in the subject group and reduced the strength of the results. Some subjects had only cortical damage while others had damage to their subcortical structures as well. As we have discussed above, this can have an undesirable influence on the results and make it very difficult to determine what influence each region of the brain has on emotion recognition.

A related concern has to do with the reuse of research participants. Because of the limited availability of brain-injured patients, the same people often participate in many different studies. This is both good and bad. It provides an experimental control for the extent and specific nature of the brain injury, which is desirable because it helps in linking structure to function. However, reuse of subjects limits generalizability because of the fact that each participant has a unique pattern of damage. Reuse of participants also makes it difficult to reconcile inconsistent findings, which might be due to aging-related cognitive changes in the participant, as we noted above.

Another weakness of research based on brain-damaged subjects is that it is rare that researchers find subjects with the precise damage needed for their study. There usually is either some additional damage to surrounding structures (Calder et al., 1996; Hamann et al., 1996; Young, Hellawell, Van De Wal, & Johnson, 1996) or the damage to the intended organ is not complete and some of the functions of the organ remain intact (Calder et al., 1996; Young et al., 1996). A prime example is the Adolphs et al. (1999b) study. The subject, an individual with complete bilateral amygdala lesions, had no difficulties recognizing facial emotion. This finding was contradictory to the preponderance of evidence found in research of this type including other studies that have used this very same subject. Adolphs et al. (1999b) found instead of emotion recognition deficits, deficits in the subject's ability to rate the intensity of the presented emotions.

Just as the extent of brain insult can limit the research questions that can be

addressed, it is also possible for researchers to approach the study of brain-injured participants in an overly restrictive way. An example of this is when researchers focus solely on one structure, like the amygdala, in investigations of brain response to facial emotional stimuli (e.g., Irwin, Davidson, Lowe, Mock, Sorenson, & Turski, 1996). As subsequent research has demonstrated, this is not adequate to form a complete picture of neural activity in response to negative facial expressions.

As noted earlier, there are also problems introduced by treating the amygdala as a single unitary structure; in fact it has many internal subdivisions composed of distinct neural cell groupings (Heimer, 1995). When research is done on subjects with amygdala damage, it is difficult to be sure about which of the cell groupings are damaged and if, in cases of bilateral amygdala damage, both amygdalae have symmetrical damage. Along similar lines, it is debated which brain structures should be included under the blanket term "basal ganglia;" different researchers have used different criteria to determine what this term refers to, which has resulted in a heterogeneous group of opinions over many years. Some structures typically included are the caudate nucleus, putamen, and globus pallidus (pallidum). When studies report damage in the basal ganglia but do not specify what this really means in terms of real damage, the reader is at a loss to know what they should take from the results and how they should interpret them. Karow et al. (2001) for instance, used subjects with unilateral cortical-only damage in the left and right hemispheres and also used subjects with subcortical damage as well. The subcortical damage was in the basal ganglia but there was no explanation as to which brain structures were in reality damaged. This lack of specificity weakens the conclusions that can be drawn from their results.

A related concern has to do with the diagnostic label of traumatic brain injury. As noted above, traumatic brain injury encompasses many different kinds of trauma which cause damage to more than one brain structure. As should now be clear, this limits both the specificity and the generalizability of the results of subjects with traumatic brain injury.

Positron Emissions Tomography (PET) and functional magnetic resonance imaging (fMRI) are both convenient and commonly used ways to view brain activity while subjects perform specific functions. Use of these techniques, and others, with normal subjects allows for comparison to the results of studies using brain damaged subjects, and thus in principle can provide converging evidence that should help differentiate between the competing theories of emotion recognition. In our view, such comparisons to normal controls are needed in order to complete the theoretical picture of brain function. These techniques must be used with caution, however; it is important that PET scans be compared to accurate and current MRI scans to know for certain which anatomical structures show the heightened activation. There are two concerns here, when PET and MRI scans don't converge (Morris et al., 1998b), and when PET scans and behavioral data from brain-damaged subjects don't converge.

A final set of cautionary statements concerns details of the experimental procedures used. For example, Lane et al. (1997a) did not include confirmatory ratings of the emotions being studied. Without this information it is difficult to be sure that the emotions presented to participants were in fact those that the researchers intended. This is a concern not only with their emotional stimuli, but with their putatively neutral stimuli as well. The neutral stimuli were pictures of natural scenes, and it could be argued that these have a calming, pleasant, rejuvenating effect. If this is true, then it may not be accurate to characterize these as neutral.

One way of addressing the issue of stimulus adequacy is to use well-standardized stimulus items. In fact, many of the experimental papers written about facial expression recognition research have use a variant of stimuli from the Ekman and Friesen (1976) series of facial expressions. The use of standardized stimuli leads to more solid confidence in their results, although of course it can limit generalizability. There does not appear to be an analogous standardized set of auditory stimuli, however. These are generally

created by the individual research groups and, therefore, have less consistency.

The exact method of stimulus presentation also makes a difference and should be considered. Wurm, Vakoch, Strasser, Calin-Jageman and Ross (2001) performed a study in which they tested the influence of either randomized (experiment one) or blocked (experiment two) stimulus presentation. There were two conditions, an emotionally congruent condition in which the affective sentences were recorded in the same affective tone, and an incongruent condition in which the sentences were recorded in a different affective tone. They found that emotional stimuli presented in the blocked form led to facilitation of recognition; there was a significant reaction time advantage for the congruent condition. However, in the randomized trials, there was no facilitation of recognition. The researchers explain that the blocked stimuli created an expectancy which influenced the subjects' reactions to the stimuli. These results indicate that caution must be given to experiments that use blocked emotional stimulus sets. It is conceivable that brain damaged individuals may not be able to take advantage of this kind of expectancy. This might have the effect of causing researchers to overestimate the deficit in their functioning, compared to normal control subjects who can make use of stimulus-based expectancies.

We should also note, following Calder et al. (1996), the importance of establishing face recognition abilities in their brain damaged subjects. Fortunately most studies have taken care to do this. However, some have not, and conclusions about deficiencies in face *emotion* recognition are tenuous at best if the participants are also deficient in recognizing facial identity in general.

Conclusion

In this paper we have reviewed the existing literature on emotion recognition and its relationship to brain structures. In this section we summarize the major points we have raised and make some suggestions about possible directions for future research.

One of the points we wish to emphasize is that the previously-assumed view of the amygdala as the primary emotion recognition center needs significant modification. A great deal of research suggests that other brain structures contribute to emotion recognition, which argues against this view. In addition, research suggests that the amygdala may play a role not only in emotion recognition, but also in the initial encoding of emotion information and analysis of emotion intensity. As the Cahill et al. (1996) study demonstrated, the amygdala is also involved in long-term memory associated with emotional stimuli.

Another long-standing view that requires revision is that the right hemisphere is the "emotion hemisphere." Studies looking at people with damage to one or the other hemisphere have demonstrated that the left hemisphere can contribute to emotion recognition, and that the left hemisphere in fact has the ability to recognize emotion.

In our view, the preponderance of evidence suggests that many brain structures function together to accomplish emotion recognition. Researchers have found that a given emotion activates a unique set of structures, but also that a given structure becomes activated in response to more than one emotion. That is, there appears to be substantial overlap in structure and function which is not conveniently defined by hemispheric boundaries. Furthermore, the valence theory, which was once accepted as an accurate description of emotion recognition, falls short. As we have discussed, activation of these sets of brain structures is not dependent on the valence of the emotion stimuli.

The method by which emotional stimuli are presented to research participants also matters. Most studies in this area present stimuli visually, but there have been some that used auditory presentation. There are some obvious differences in the results of these kinds of studies, owing to the fact that different brain regions are responsible for visual vs. auditory processing. For example, auditory processing of emotion is difficult for people with temporal lobe damage, but this makes sense given that auditory processing in general is handled in the temporal lobe. Conclusions about emotion recognition for the two perceptual modalities remain speculative at this point, because of this link between

specific brain regions and specific perceptual modalities and also because of the relatively small number of studies using auditory stimuli.

It would be beneficial to have additional studies using auditory stimulus presentation for another reason, too. The emotional aspect of spoken language (prosody) helps individuals with very basic aspects of perception (e.g., breaking the speech signal into individual words), as well as higher-level communicative functions (e.g., resolving ambiguity, correctly interpreting sarcasm, and understanding indirect requests). Emotion is important for arriving at an accurate understanding of verbal messages and situations. Walker, Fongemie, and Daigle (2001) demonstrated that prosody affects the extent to which sentences are judged to be natural-sounding, for both normal controls and brain-damaged participants, but there has not been a great deal of research looking at the precise role of various brain structures in the use of prosodic information.

Additional empirical work on the extent to which normal subjects can control facial expressions and subtle vocal characteristics would be very useful. People often wish to conceal their true feelings or beliefs (whether in situations involving deliberate deception, or in more innocuous situations involving personal opinions), and there are often subtle cues available in prosody or facial expressions. Future research on just how much control people have over their supralinguistic messages would lead to a better theoretical understanding of how brain damage affects these processes. In addition, this research would be very helpful in furthering our understanding of how attempted deception can be detected, in both formal (e.g., criminal) and everyday settings.

In this paper we have attempted to integrate recent neuropsychological findings in the area of emotion processing. Sympathetic autonomic arousal and emotion are central to traditional methods of lie detection, so it is reasonable to suppose that people with interests in this area can benefit from a knowledge of the link between brain activity and emotion. In addition, by extending current investigations into adjacent topic areas, and by making use of the increasingly sophisticated imaging techniques likely to become available in the future, researchers can work toward a more complete understanding of the relationships between emotion, deception, and the structure and function of the brain.

Received August 1, 2002
Revision received February 10, 2003
Accepted May 7, 2003

References

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A.R. (1998a). Cortical systems for the recognition of emotion in facial expressions. *The Journal of Neuroscience, 16,* 7678-7687.

Adolphs, R., Russell, J.A., & Tranel, D. (1999b). A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. *Psychological Science, 10,* 167-171.

Adolphs, R., Schul, R., & Tranel, D. (1998b). Intact recognition of facial emotion in Parkinson's Disease. *Neuropsychology, 12,* 253-258.

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia, 37,* 1285-1292.

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology Special Issue, 15,* 396-404.

Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372,* 669-672.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A.R. (1995). Fear and the human amygdala, *Journal of Neuroscience, 15,* 5879-5891.

Adolphs, R., Tranel, D., Hamann, S., Young, A.W., Calder, A.J., Phelps, E.A., Anderson, A., Lee, G.P., & Damasio, A.R. (1999a). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37,* 1111-1117.

Anderson, A.K. & Phelps, E.A. (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *NeuroReport, 9,* 3607-3613.

Anderson, A.K., & Phelps, E.A. (2000). Expression without recognition: Contributions of the human amygdala to emotional communication. *Psychological Science, 11,* 106-111.

Anderson, A.K., Spencer, D.D., Fulbright, R.K., & Phelps, E.A. (2000). Contribution of the anteromedial temporal lobes to the evaluation of facial emotion. *Neuropsychology, 14,* 526-536.

Baum, S.R., Pell, M.D., Leonard, C.L., & Gordan, J.K. (1997). The ability of right and left hemisphere damaged individuals to produce and interpret prosodic cues marking phrasal boundaries. *Language and Speech, 40,* 313-330.

Blair, R.J.R., Morris, J.S., Frith, C.D., Perrett, D.I., & Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain, 122,* 883-893.

Borod, J.C., Cicero, B.A., Obler, L.K., Welkowitz, J., Erhan, H.M., Santschi, C., Grunwald, I.S., Agosti, R.M., & Whalen, J.R. (1998). Right Hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12,* 446-458.

Buller, D.B., & Burgoon, J.K. (1998). Emotional expression in the deception process. In P.A. Andersen & L.K. Guerrero (Eds.), *Handbook of communication and emotion: Research, theory, applications, and contexts* (pp. 381-402). New York: Academic Press.

Cahill, L., Haier, R.J., Fallon, J., Alkire, M.T., Tang, C., Keator, D., Wu, J. & McGaugh, J.L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Science USA, 93,* 8016-8021.

Calder, A.J., Keane, J., Manes, F., Antoun, N., & Young, A.W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience, 3,* 1077-1078.

Calder, A.J., Young, A.W., Rowland, D., Perrett, D.I., Hodges, J.R., & Etcoff, N.L. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13,* 699-745.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11,* 86-89.

Ekman, P. (2001a). Detecting deceit from words, voice, or body. In P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (3rd edition, pp. 80-122). New York: W.W. Norton & Company.

Ekman, P. (2001b). Why lies fail. In P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (3rd edition, pp. 43-79). New York: W.W. Norton & Company.

Ekman, P., & Friesen, W.V. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists.

Ghika-Schmid, F., Ghika, J., Vuilleumier, P., Assal, G., Vuadens, P., Scherer, K., Maeder, P., Uske, A., & Bogousslavsky, J. (1997). Bihippocampal damage with emotional dysfunction: Impaired auditory recognition of fear. *European Neurology, 38,* 276-283.

Hamann, S.B., Ely, T.D., Hoffman, J.M., & Kilts, C.D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science, 13,* 135-141.

Hamann, S.B., Stefanacci, L., Squire, L.R., Adolphs, R., Tranel, D., Damasio, H., & Damasio, A.R. (1996). Recognizing facial emotion. *Nature, 379,* 497.

Heimer, L. (1995). *The human brain and spinal cord: Functional neuroanatomy and dissection guide* (2nd ed.). New York: Springer-Verlag.

Irwin, W., Davidson, R.J., Lowe, M.J., Mock, B.J., Sorenson, J.A., & Turski, P.A. (1996). Human amygdala activation detected with echo-planar functional magnetic resonance imaging. *NeuroReport, 7,* 1765-1769.

Jackson, J.H. (1874). On the nature of the duality of the brain. *The Medical Press and Circular, 1,* 41-44.

Jackson, J.H. (1880). On affections of speech from disease of the brain. *Brain, 2,* 203-222.

Karow, C.M., Marquardt, T.P., & Marshall, R.C. (2001). Affective processing in left and right hemisphere brain-damaged subjects with and without subcortical involvement. *Aphasiology, 15,* 715-729.

Lane, R.D., Reiman, E.M., Ahern, G.L., Schwartz, G.E., & Davidson, R.L. (1997a). Neuroanatomical correlates of happiness, sadness, and disgust. *American Journal of Psychiatry, 154,* 926-933.

Lane, R.D., Reiman, E.M., Bradley, M.M., Lang, P.J., Ahern, G.L., Davidson, R.J. & Schwartz, G.E. (1997b). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia, 35,* 1437-1444.

Langleben, D.D., Schroeder, L., Maldjian, J.A., Gur, R.C., McDonald, S., Ragland, J.D., O'Brien, C.P., & Childress, A.R. (2002). Rapid communication: Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage, 15,* 727-732.

Ley, R.G., & Bryden, M.P. (1979). Hemispheric differences in processing emotions and faces. *Brain and Language, 7,* 127-138.

Lorch, M.P., Borod, J.C., & Koff, E. (1998). The role of emotion in the linguistic and pragmatic aspects of aphasic performance. *Journal of Neurolinguistics, 11,* 103-118.

Mandal, M.K., Tandon, S.C., & Asthana, H.S. (1991). Right brain damage impairs recognition of negative emotions. *Cortex, 27,* 247-253.

Marquardt, T.R., Rios-Brown, M., Richburg, T., Seibert, L.K., & Cannito, M.P. (2001). Comprehension and expression of affective sentences in traumatic brain injury. *Aphasiology, 15,* 1091-1101.

Mills, C.K. (1912a). The cerebral mechanism of emotional expression. *Transactions of the College of Physicians of Philadelphia, 34,* 381-390.

Mills, C.K. (1912b). The cortical representation of emotion, with a discussion of some points in the general nervous mechanism of expression in its relation to organic nervous disease and insanity. *Proceedings of the American Medico-Psychological Association, 19,* 297-300.

Morris, J.S., Friston, K.J., Büchel, C., Frith, C.D., Young, A.W., Calder, A.J., & Dolan, R.J. (1998b). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain, 121,* 47-57.

Morris, J.S., Friston K.J., & Dolan, R.J. (1997). Neural responses to salient visual stimuli. *Proceedings of the Royal Society London B: Biological Science, 264,* 769-775.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., & Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812-815.

Morris, J.S., Öhman, A., & Dolan, R.J. (1998a). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393,* 467-470.

Morris, J.S., Öhman, A., & Dolan, R.J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Science USA, 96,* 1680-1685.

Papez, J.W. (1937). A proposed mechanism of emotion. *Archives of Neurology and Psychiatry, 38,* 725-743.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C.R., Bullmore, E.T., Brammer, M., & Gray, J.A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society London B: Biological Science, 265,* 1809-1817.

Pollina, D.A., Vakoch, D.A., & Wurm, L.H. (1998). Formant structure of vowels spoken during attempted deception. *Polygraph, 27,* 96-107.

Reuter-Lorenz, P., & Davidson, R.J. (1981). Differential contributions of the two cerebral hemispheres for perception of happy and sad faces. *Neuropsychologia, 19,* 609-614.

Reuter-Lorenz, P., Givis, R.P., & Moscovitch, M. (1983). Hemispheric specialization and the perception of emotion: Evidence from right-handers and from inverted and non-inverted left-handers. *Neuropsychologia, 21,* 687-692.

Scott, S.K., Young, A.W., Calder, A.J., Hellawell, D.J., Aggleton, J.P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature, 385,* 254-257.

Spell, L.A., & Frank, E. (2000). Recognition of nonverbal communication of affect following traumatic brain injury. *Journal of Nonverbal Behavior, 24,* 285-300.

Spence, S.A. Farrow T.F., Herford, A.E., Wilkinson, I.D., Zheng, Y., & Woodruff, P.W. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport, 12,* 2849-2853.

Sprengelmeyer, R., Rausch, M., Eysel, U.T. & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society London B: Biological Science, 265,* 1927-1931.

Stone, V.E., Nisenson, L., Eliassen, J.C., & Gazzaniga, M.S. (1996). Left hemisphere representations of emotional facial expressions. *Neuropsychologia, 34,* 23-29.

Vrij, A. (2000). Non-verbal behaviour during deception. In A. Vrij, *Detecting lies and deceit: The psychology of lying and the implications for professional practice* (pp. 21-55). New York: John Wiley & Sons.

Walker, J.P., Fongemie, K., & Daigle, T. (2001). Prosodic facilitation in the resolution of syntactic ambiguities in subjects with left and right hemisphere damage. *Brain and Language, 78,* 169-196.

Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., & Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18,* 411-418.

Wurm, L.H., Vakoch, D.A., Strasser, M.R., Calin-Jageman, R., & Ross, S.E. (2001). Speech perception and vocal expression of emotion. *Cognition and Emotion, 15,* 831-852.

Young, A.W., Hellawell, D.J., Van De Wal, C. & Johnson, M. (1996). Facial expression processing after amygdalotomy. *Neuropsychologia, 34,* 31-39.

Relative Efficacy of the Utah, Backster, and Federal Scoring Rules: A Preliminary Investigation

Donald J. Krapohl & Andrew B. Dollins[1]

Abstract

This report describes a rudimentary comparison of the three major 7-position systems used to evaluate physiological data collected during a polygraph examination. The Utah, Backster, and Federal scoring rules were evaluated by comparing values assigned to chart segments of 100 polygraph examinations by field examiners who blind scored the physiological data using one of the three scoring systems. Values assigned by examiners using all three systems distinguished between deceptive and truthful examinees. Statistically significant differences were found between the scores assigned using specific systems suggesting differences among the scoring rules. While intra-examiner reliability was high for all scorers, some significant differences were found between examiners using a specific system. The patterns of scores obtained across scoring systems, and among examiners using the same scoring system, are described. Greater inter-scorer differences were found among examiners using the most complex, Backster, scoring rules and the least inter-scorer differences were found among examiners using the least complex, Utah, scoring rules. It is further suggested that symmetrical decision criteria may not be optimal for all three scoring systems.

Introduction

The numerical analysis of polygraph charts introduced by Cleve Backster (1962) is considered by many to be one of the great advances in the field of polygraphy (Raskin & Honts, 2002). Prior to Backster's introduction of numerical analysis, polygraph examiners used a global approach to chart evaluation, an approach fraught with subjectivity and idiosyncrasies. Backster's 7-position concept provided examiners, for the first time, with a method for objectively quantifying responses, and establishing uniform decision rules. Seven-position scoring has been widely adopted, and Backster provided the foundation for two other similar systems, the Federal and the Utah scoring systems. These three systems currently dominate field practice in the U.S.

Though there are many adherents to each of these three systems, there is little data to discern which of them may produce the best accuracies. We were only able to locate one paper that directly compared all three scoring systems (Weaver, 1985), but it shed little light on the question. The study used only 15 cases conducted by the author, and rescored only by the author, limitations that would seriously impair result generalizability. Moreover, ground truth was not used as the criterion against which to compare the three scoring systems in Weaver's research. Weaver only addressed the degree of decision agreement among his three scorings of the same 15 cases and concluded that the Backster, Utah, and Federal scoring systems all result in similar decisions.

All polygraph numerical scoring systems have three common components: scoring rules, computation rules, and decision rules. Scoring rules encompass those relating to selection of tracing features, artifact rejection, choice of question pairs for comparison, and number-assignment schemes. Computation rules dictate how the assigned numbers are combined and weighted.

[1]Donald Krapohl and Dr. Andrew Dollins are researchers at the U.S. Department of Defense Polygraph Institute. This paper is one in a series for the project titled Best Practices. Requests for reprints should be directed to: Donald J. Krapohl, DoD Polygraph Institute, 7540 Pickens Ave., Ft. Jackson, SC 29207, or via e-mail to: krapohld@jackson-dpi.army.mil.

Decision rules govern the relationship between computation rules and decisions of Deception Indicated (DI), No Deception Indicated (NDI), or Inconclusive. Because of their fundamental importance to the subsequent processes and the dearth of systematic comparative research, scoring rules of the Backster, Federal, and Utah scoring systems are the focus of this investigation.

It is difficult to systematically quantify the similarities and differences among the scoring rule systems because of differences in notation, the paucity of detailed published descriptions, and changes over time. Some general differences among the systems, however, warrant comment. The Backster Zone Comparison Technique (ZCT), taught at the Backster School of Lie Detection has, according to Capps and Ansley (1992), undergone major modifications in 1979 and 1984, but has been relatively stable since 1990 (refresher lectures have been presented annually at American Polygraph Association seminars since 1996). While the Backster system requires examiners to evaluate a conservative eight to ten distinct physiological reactions, the subsequent rules for assigning values to reactions are undoubtedly the most complex of the three systems. Prior to attempting to translate the 21 Backster system rules into more common polygraph terms, Capps and Ansley (1992, p. 291) state that "The Backster rules are written so that only those trained by Backster, or those familiar with his terminology can understand them." Backster has not released his system beyond his school, except in summary form, and there has been no systematic examination (e.g., occurrence frequency in or proportion of accuracy accounted for) of these rules published.

An instruction manual on chart interpretation dated October 1962 and used by the United States Army Military Police School (renamed the Department of Defense Polygraph Institute, DoDPI, in 1986) relied heavily on guidelines provided by Cleve Backster for his Zone Comparison Technique. In more recent times, Swinford (1999) described the DoDPI numerical evaluation system as composed of 23 physiological criteria with a few relatively simple guidelines for assigning values to responses. The most current Federal Forensic Psychophysiology Program Test Data Analysis guidelines (DoDPI, February, 2003) include 20 primary diagnostic features and three secondary considerations, still with relatively simple guidelines for assigning values to responses. While the Federal system has more potential scoring features than the other two systems, research has shown that practitioners actually employ a common subset of those that are taught (Capps & Ansley, 1992). Blackwell (1998) found that the Federal scoring system produced good inter-scorer agreement (*kappa* = .57), and decision accuracy of about 89%.

The Utah scoring rules are the least complex of the systems examined. The 10 physiological criteria formally described by Bell, Raskin, Honts, and Kircher (1999) differ little from those described by Raskin, Barland, and Podlesny in 1977. It should be noted that two of the Utah physiological criteria are for a photoplethysmograph, a fifth channel not used by the Backster or Federal systems. The Utah scoring rules, which were developed and refined through systematic investigation, have repeatedly produced accuracies at or above 90% with 95% agreement among blind scorers (Raskin & Honts, 2002).

While the three numerical scoring systems stem from a common origin, they differ markedly in their current forms. Each has, however, been used successfully for a number of years. This project was undertaken to determine if there are practical differences among the scoring rules. It is hypothesized that points assigned by examiners using the scoring rule systems will not differ significantly in accuracy or reliability.

Method

Polygraph Cases

One hundred polygraph cases were randomly drawn from the database at the DoDPI with the constraint that half of the cases were of deceptive examinees, and the remaining cases were of truthful examinees. All 100 cases were of criminal suspects, and ground truth had been established by a confession that inculpated or exculpated the examinee, or by the discovery of other highly reliable forensic evidence. All were recorded with Axciton computer polygraphs (Axciton

Systems, Inc., Houston, TX) by local, state, or federal law enforcement agencies. All polygraph cases were three-question single-issue examinations in which the DoDPI ZCT (DoDPI, 1992) was used and a minimum of three charts were recorded. The question sequence for the DoDPI ZCT is summarized in Table 1.

Table 1
Department of Defense Polygraph Institute Zone Comparison Technique Question Summary

Question Number	Question Type
1	Irrelevant
2	Sacrifice Relevant
3	Symptomatic
4 C	Comparison
5 R	Relevant
6 C	Comparison
7 R	Relevant
8	Symptomatic
9 C	Comparison
10 R	Relevant

Chart Preparation

Segments of the polygraph charts from the 100 cases were photocopied. All segments contained questions 4C (comparison question), 5R (first relevant question), and 6C (comparison question). These portions of the charts were selected to test unique scoring rules that require bracketed relevant questions. In the DoDPI ZCT, only 5R is bracketed in this manner.

The 100 3-chart cases resulted in 300 photocopied segments. From the 100 cases, all three spots from the first 10 deceptive and the first 10 truthful cases were photocopied twice. These duplicate segments were used to provide a measure of intra-scorer reliability. All photocopied segments were devoid of identifying information such as case numbers, file numbers, and examination dates. Addition of the repeated segments created a total of 360 segments. The order of the 360 segments was randomized with the limitation that the originals of the 60 repeated segments were among the first 120 segments and the 60 repeated segments were among the last 120 segments. The same segments were used with all scorers, and in the same order.

Scorers

Thirteen scorers were recruited through an announcement in the official newsletter (APA Newsletter, 2001, 34(5), 33) of the American Polygraph Association. Of the 13 scorers, four were graduates of the Backster School of Lie Detection, and six were graduates of the DoDPI and in federal or local service. Of the three Utah scorers, one had attended an advanced training course taught by Dr. David Raskin, then of the University of Utah. Two Utah scorers were DoDPI graduates who had no field experience at that time. They were instructed to use the article on polygraph scoring by Bell, Raskin, Honts, and Kircher (1999) as their only scoring guide.

All scorers were made aware that this was a study designed to compare the three major scoring systems. None were informed of ground truth, base rates, or that some segments they received were duplicates of earlier segments, until all of the data were collected.

Procedure

Three packets were prepared for each scorer. Each packet contained: the study instructions (Appendix A), score sheets (Appendix B), and one-third of the 360 photocopied chart segments (Example in Appendix C). Each packet was mailed or hand delivered to the scorers. Release of the second and third packets was contingent on return of the completed previous packet. Scorers were instructed to employ the 7-position scoring system according to the rules they had been taught, and not to share any information with other participants until the study was completed. If the data contained artifacts which precluded a score assignment, examiners were instructed to mark an "N" on the score sheet space instead of a score.

Data Reduction and Analysis

The hand-written score sheet data were tabulated in Microsoft Excel and reordered to their original cases. The scores were then summed by case and these sums were the dependent measure. A .05 significance criterion was used for all inferential statistical

tests. The Greenhouse-Geisser correction for violation of the sphericity assumption was applied, where appropriate, to the repeated measure analysis of variance (ANOVA) calculations described in the Results section. The Bonferroni procedure was used to adjust significance criteria to compensate for multiple comparisons (Rosenthal & Rosnow, 1991) when pair-wise *t*-tests were calculated. Eta squared (η^2) is provided as a measure of effect size (Rosenthal & Rosnow, 1991). For *F* tests with 1 numerator degree of freedom, η is equivalent to Pearson's product-moment correlation, *r*. That is, when two groups are compared, r^2 (and η^2), are the proportion of total variance accounted for by differences between the two groups.

Results

Artifacts

Table 2 shows the frequency that artifacts were identified by individual examiners. The average percentages of artifact rejection for the 900 comparisons (300 segments times 3 channels) were 6.9%, 4.5%, and 7.5% for the Utah, Backster, and Federal examiners, respectively. Tests of proportions for these rates indicated no significant difference between scores assigned by the Utah and Federal examiners ($z = 0.91$, $p > .05$, $\eta^2 = .00$), but a significant difference was found between scores assigned by the Utah and Backster examiners ($z = 4.17$, $p < .05$, $\eta^2 = .00$), and between the Backster and Federal examiners ($z = 5.71$, $p < .05$, $\eta^2 = .00$). While the significant differences suggest that the Backster examiners had a lower artifact rejection rate than either the Utah or Federal examiners, the difference was small and accounted for less than 1% of the total comparison variance.

Because the sample size was fixed, a decision that a tracing contained an artifact incrementally reduced the opportunity for an examiner to assign a score because that tracing was not evaluated. Omitting tracings with artifacts could have improved accuracy because misleading information was excluded, or reduced accuracy because potentially useful

Table 2
Frequency and Proportions of Artifact Decisions

Evaluation System	Scorer	Channel			Total	Average Proportion
		Pneumograph	EDA	Blood Volume		
Utah	1	32	26	24	82	.09
	2	8	25	12	45	.05
	3	24	19	17	60	.07
Backster	1	1	11	1	13	.01
	2	83	13	7	103	.11
	3	16	10	5	31	.03
	4	4	6	5	15	.02
Federal	1	18	7	5	30	.03
	2	46	14	17	77	.09
	3	4	19	41	64	.07
	4	85	24	31	140	.16
	5	70	12	11	93	.10
	6	0	0	0	0	.00

information was ignored. A test for a relationship between the number of artifacts noted by an examiner, and the number of times the total scores were on the correct side of zero produced a weak correlation (r = 0.23) which was not significant, $t[98] = 0.73$, $p > .05$, $\eta^2 = .00$. This suggests that, within the limits of these data, artifact decisions did not influence decision accuracy among the scoring systems.

Intra-scorer Reliability

Pearson product-moment correlations (r) and correlated t-tests were calculated for the scores assigned to 20 original cases and their duplicates. Within-scorer reliability was uniformly high, as seen in Table 3. Correlations for individual scorers were all significantly different from chance ($p < .05$). None of the t-test results were significantly different from 0 ($p > .05$) and differences between means accounted for less than 10% of the variance.

Table 3
Intra-scorer Pearson Product-Moment Correlation Coefficients (r), Correlated t-test Results, and Effect Sizes (η^2) Evaluating Scores Assigned to 20 Original and Duplicate Cases

Evaluation System	Scorer	r	$t(19)$	η^2
Utah	1	.93 *	.32	.00
	2	.93 *	.41	.01
	3	.93 *	1.34	.09
	Average	.93		
Backster	1	.95 *	1.04	.05
	2	.95 *	.41	.01
	3	.80 *	-.88	.04
	4	.96 *	-.54	.02
	Average	.91		
Federal	1	.91 *	-.12	.00
	2	.99 *	.94	.04
	3	.95 *	.00	.00
	4	.96 *	-.38	.01
	5	.93 *	-.31	.00
	6	.90 *	-.91	.04
	Average	.94 *		

* $p < .05$)

Inter-scorer Reliability

Inter-scorer reliability was high, as indicated by the Pearson Product-Moment correlation coefficients in Table 4—all of which are significantly better than would be expected by chance alone. Results of a 2-group (deceptive versus truthful) by 13-examiner ANOVA indicate that values assigned to data from deceptive individuals were, on average, 6.11 points lower than those assigned to data from truthful individuals, $F(1, 98) = 75.99$, $p < .05$, $\eta^2 = .44$. There was a significant difference among points assigned by examiners, $F(12, 1176) = 19.92$, $p < .05$, $\eta^2 = .17$, indicating that some examiners differed from other examiners when evaluating the same data. The examiner by group interaction was also significant, $F(12, 1176) = 6.13$, $p < .05$, $\eta^2 =$

Table 4
Inter-scorer Pearson Product-Moment Correlation Coefficients (r) for Examiners Using the Same Test Data Evaluation System

Evaluation System	Examiners Compared	Examiners Compared 2	3	4	5	6
		Truthful and Deceptive Combined (df = 99)				
Utah	1	.87	.87			
	2		.87			
Backster	1	.86	.81	.83		
	2		.71	.80		
	3			.76		
Federal	1	.78	.69	.78	.80	.78
	2		.85	.93	.91	.86
	3			.85	.81	.79
	4				.88	.84
	5					.86
		Deceptive (df = 49)				
Utah	1	.84	.85			
	2		.85			
Backster	1	.78	.79	.84		
	2		.63	.83		
	3			.80		
Federal	1	.72	.60	.73	.76	.69
	2		.84	.94	.92	.80
	3			.81	.78	.72
	4				.88	.82
	5					.81
		Truthful (df = 49)				
Utah	1	.73	.73			
	2		.70			
Backster	1	.72	.58	.67		
	2		.44	.58		
	3			.53		
Federal	1	.61	.53	.58	.63	.66
	2		.72	.80	.77	.75
	3			.75	.65	.67
	4				.70	.64
	5					.73

Note. $p < .05$ for all correlations.

.06, indicating that some examiners evaluated data from deceptive and truthful individuals differently than other examiners. Contrasts indicate that when evaluating the same data, points assigned by examiners using the Utah system were significantly different from those assigned by examiners using the Backster system, $F(1, 98) = 5.38$, $p < .05$, $\eta^2 = .05$; points assigned by examiners using the Utah system were significantly different from those assigned by Federal system examiners, $F(1, 98) = 30.44$, $p < .05$, $\eta^2 = .24$; and points assigned by examiners using the Backster system were significantly different from those assigned using the Federal system, $F(1, 98) = 39.50$, $p < .05$, $\eta^2 = .29$. In sum, each of the scoring systems produced total scores that were different from the other two.

To assist in result interpretation, separate analyses were calculated for points assigned to data collected from truthful and deceptive individuals. These data are plotted in Figure 1. For data from deceptive individuals, contrasts indicate that on average, evaluations using the Utah system were 1.12 points higher than those assigned using the Backster system, $F(1, 49) = 9.02$, $p < .05$, $\eta^2 = .15$; evaluations using the Utah system were 1.63 points lower than those assigned using the Federal system, $F(1, 49) = 30.77$, $p < .05$, $\eta^2 = .38$; and evaluations using the Backster system were 2.75 points lower than those assigned using the Federal system $F(1, 49) = 18.97$, $p < .05$, $\eta^2 = .28$. For data from truthful individuals, contrasts indicate that on average, evaluations using the Utah system were .14 points higher than, but not significantly different from, those assigned using the Backster system, $F(1, 49) = .12$, $p > .05$, $\eta^2 = .00$; evaluations using the Utah system were .60 points lower than those assigned using the Federal system, $F(1, 49) = 4.68$, $p < .05$, $\eta^2 = .09$; and evaluations using the Backster system were .73 points lower than those assigned using the Federal system, $F(1, 49) = 77.14$, $p < .05$, $\eta^2 = .61$. These analyses suggest that, with the single exception of truthful individual data evaluated using the Backster and Utah systems, there are significant differences among the points assigned to the same data using different evaluation systems.

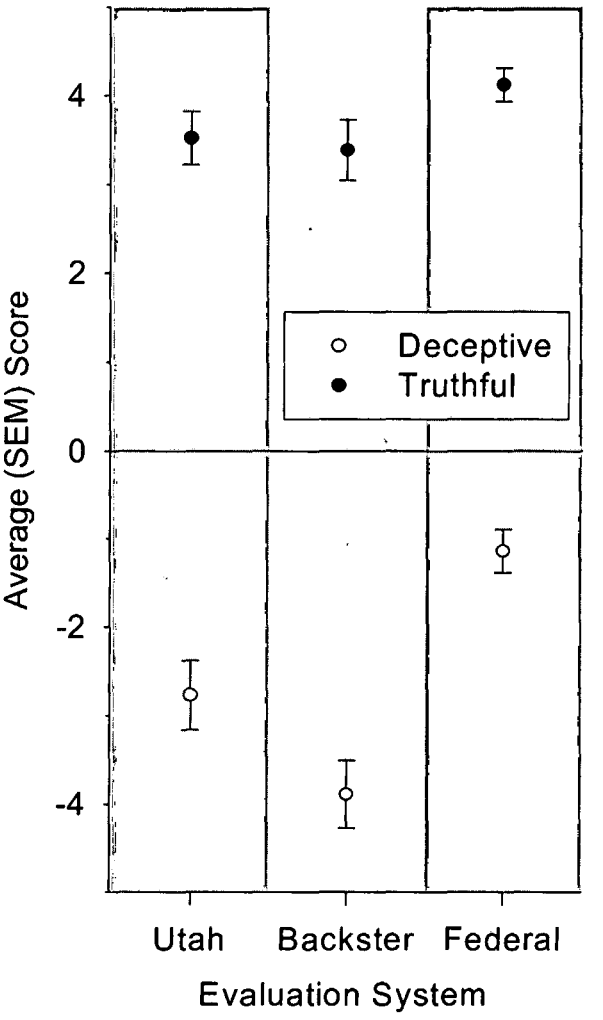


Figure 1. Average (*SEM*) points assigned, by evaluation system group, when examiners evaluated data from 50 deceptive and 50 truthful examinees.

The points assigned by examiners using the Utah, Backster, and Federal evaluation systems were analyzed separately to better understand evaluation system characteristics. Statistically significant deceptive versus truthful data group differences were found among examiners using the Utah, $F(1, 98) = 65.36$, $p < .05$, $\eta^2 = .40$; Backster, $F(1, 98) = 72.24$, $p < .05$, $\eta^2 = .42$, and Federal, $F(1, 98) = 61.05$, $p < .05$, $\eta^2 = .38$, evaluation systems, suggesting that examiners

using all three systems evaluated data from deceptive and truthful individuals differently. Examiners using the Backster evaluation system differed significantly among themselves (i.e., examiner effect) in scores assigned to the same data, $F(3, 294) = 30.13$, $p < .05$, $\eta^2 = .24$, while examiners using the Federal and Utah evaluation systems did not [i.e., $F(5, 490) = 1.38$, $p > .05$, $\eta^2 = .02$ and $F(2, 196) = 2.57$, $p > .05$, $\eta^2 = .02$ respectively]. This suggests that there was variability in score assignment by Backster examiners that was not found among Federal and Utah examiners. Significant group

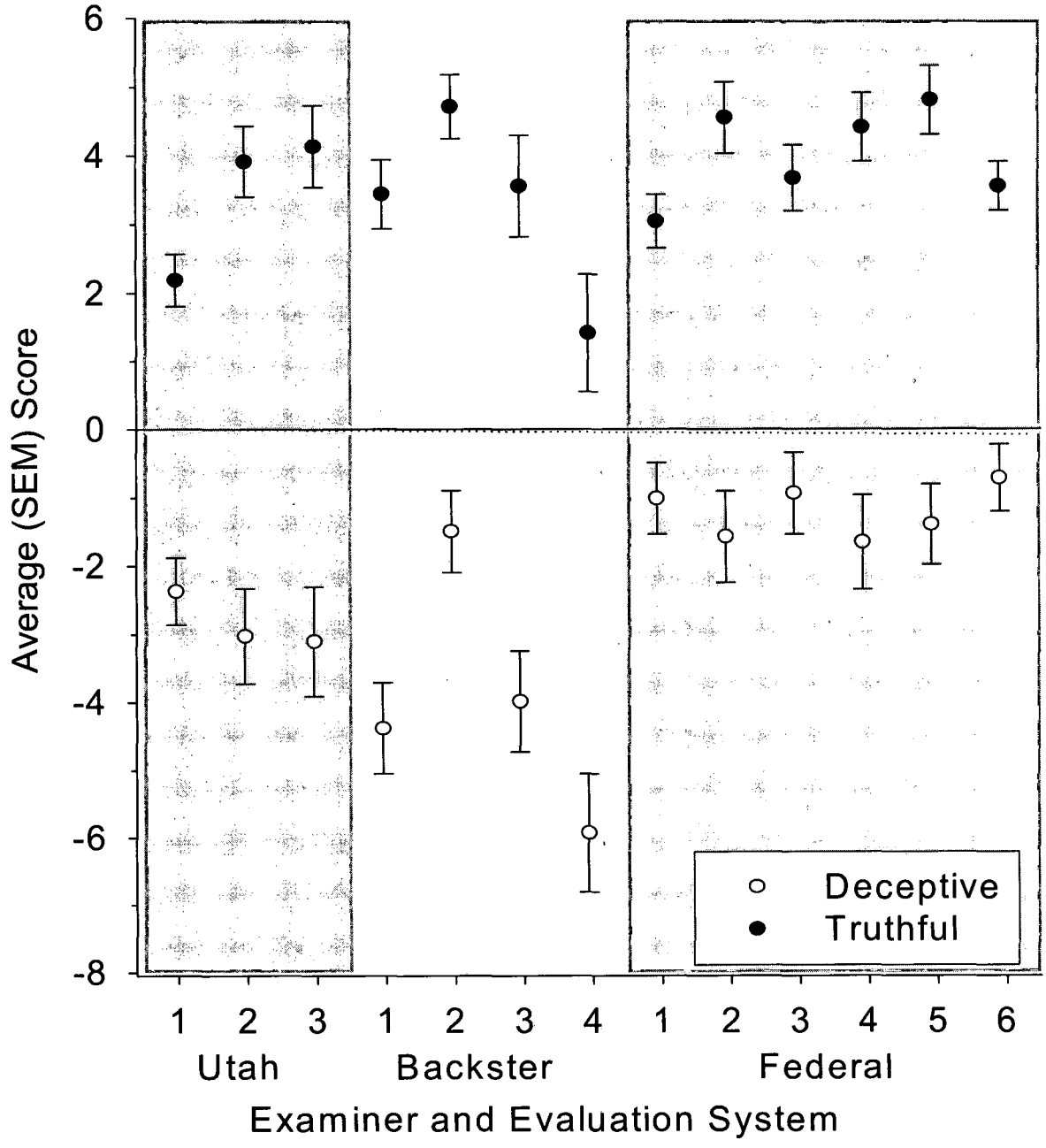


Figure 2 Average (SEM) points assigned by examiners when evaluating data from 50 deceptive and 50 truthful examinees.

(i.e., truthful versus deceptive) by examiner interactions were found among the Utah, $F(2, 196) = 54.75$, $p < .05$, $\eta^2 = .11$, and Federal, $F(5, 490) = 6.95$, $p < .05$, $\eta^2 = .06$, examiners but not among the Backster examiners, $F(3, 294) = 1.50$, $p > .05$, $\eta^2 = .02$. These interaction effects indicate that some examiners using the Federal and Utah evaluation systems were more conservative than others when assigning points to data from deceptive and truthful individuals, while no such differences were found among examiners using the Backster system. These data are illustrated in Figure 2. For example, the means of the first examiner using the Utah evaluation system are closer to each other than the means of the second and third examiners. A similar relationship can be seen for the first, third, and sixth examiners using the Federal evaluation system, but not among examiners using the Backster evaluation system.

Pairwise correlated group *t*-tests were calculated to determine if there were differences between examiners using the same evaluation system. For example, the scores of an examiner using the Federal system to evaluate data from deceptive individuals were compared to the scores of other examiners using the Federal system to evaluate the same data. Significant differences among examiners are interpreted as suggesting that examiners were not consistently applying the evaluation system rules. These data are presented in Figure 2. To conserve space, only effect sizes and significance results from the analyses are presented in Table 5 (detailed information is available on request).

There were no statistically significant differences among the scores of examiners using the Utah or Federal systems to evaluate data from deceptive individuals. Scores assigned to examinations of truthful individuals by Utah examiner one were significantly lower than those assigned by Utah system examiners two and three. All examiners using the Backster evaluation system assigned scores significantly different from the others, when evaluating deceptive individual data, except examiners one and three. When evaluating data from truthful individuals using the Backster evaluation system, examiner one assigned scores that were significantly different from those assigned by examiners two and four, and examiner two assigned scores that were significantly different from those assigned by examiner four. When evaluating data from truthful individuals using the Federal system, examiner one assigned scores that were significantly different from those assigned by examiners two, four, and five, and examiner five assigned scores that were significantly different from those assigned by examiner six.

Discussion

These data suggest five important implications for the Federal, Backster, and Utah scoring rules. First, all examiners had high score, re-score, correlation coefficients. This high intra-examiner reliability suggests that examiners were attending to their task and that they were consistently applying the scoring rules they were using. High self-agreement was required before meaningful analyses of inter-scorer and inter-system reliability could be undertaken. Whatever procedures the scorers were using in this study, they did reliability.

Second, there were statistically significant differences between scores assigned, by all examiners, to data from deceptive and truthful individuals. Given that there was high intra-examiner reliability, the Federal, Backster, and Utah scoring rules were equally effective in classifying the deceptive and truthful responses used in this study.

Third, although examiners using the three scoring rule systems differed significantly in artifact identification, that difference accounted for less than one percent of the variability. The small effect sizes suggest that the significant difference was a product of the number of observations rather than meaningful differences in artifact identification. Analyses indicate that the artifact identification rules used by examiners in this study did not influence the relative decision accuracy among the scoring systems.

Fourth, the values assigned using the three scoring rule systems, on average, differed from one another. These differences were most apparent when deceptive data were

Table 5
Effect Sizes (η^2) of Correlated Group t-tests Between Examiners Using the Same Test Data Evaluation System

Evaluation System	Examiners Compared	Examiners Compared 2	3	4	5	6
		Truthful and Deceptive Combined (df = 99)				
Utah	1	.03	.03			
	2		.00			
Backster	1	.34*	.00	.17*		
	2		.14*	.44*		
	3			.16*		
Federal	1	.02	.01	.01	.05	.03
	2		.00	.00	.01	.00
	3			.00	.01	.00
	4				.02	.00
	5					.01
		Deceptive (df = 49)				
Utah	1	.05	.05			
	2		.02			
Backster	1	.49*	.02	.18*		
	2		.27*	.61*		
	3			.21*		
Federal	1	.03	.00	.04	.02	.01
	2		.06	.00	.01	.08
	3			.06	.03	.01
	4				.01	.10
	5					.07
		Truthful (df = 49)				
Utah	1	.33*	.31*			
	2		.00			
Backster	1	.12*	.00	.06		
	2		.17*	.31*		
	3			.13*		
Federal	1	.21*	.04	.18*	.29*	.05
	2		.10	.00	.01	.15
	3			.08	.13	.02
	4				.00	.09
	5					.11*

*$p < .05$

evaluated and all three systems differed from each other. The differences were less apparent when truthful data were evaluated, and only values assigned using the Federal scoring rules differed from those assigned obtained using the Utah and Backster rules. Examination of Figure 1 suggests that values assigned using the Utah scoring rules are more symmetrical around 0 than those assigned using the Backster and Federal systems. Examiners using the Backster scoring rules assigned less positive values and examiners using the Federal scoring rules assigned more positive values to the same data.

The positive shift in scores assigned by examiners using the Federal system, for both deceptive and truthful cases, may have been due to the data used. When a relevant question is bracketed by comparison questions, the Federal scoring system requires that the comparison question eliciting the stronger response be used for scoring. This scoring rule would move all scores in a positive direction. The DoDPI ZCT has one relevant question bounded by comparison questions, and two other relevant questions that are only preceded by a comparison question. The current data did not permit an assessment of the Federal scoring system on the other two relevant questions. Though the Backster system also has rules for bracketed comparison questions, the selection of which comparison question to use depends on responses to the relevant question. Therefore, in contrast to the Federal system where the stronger comparison question is always used, scores in the Backster system could have been produced by the use of either of the comparison questions. This difference may explain why the Federal scores, and not the Backster scores, were positively shifted.

Finally, the most notable examiner performance difference is the inter-examiner consistency by scoring rule system illustrated in Figure 2. While there were between-scoring rule system differences, there were no significant differences in values assigned to data from deceptive individuals using the Utah, or using the Federal systems. There were significant differences in two of three, and four of fifteen, comparisons for examiners using the Utah and Federal scoring rules, respectively, to evaluate data from truthful individuals. Evaluating the same data, there were significant differences in five and four of six comparisons for examiners using the Backster scoring rules to evaluate data from deceptive and truthful individuals, respectively. This suggests that there was less inter-examiner consistency among examiners using the Backster scoring rules than among examiners using the Federal and Utah rules.

The group veracity by examiner interactions found among the Utah and Federal examiner scores, but not the Backster examiner scores, also have implications for those comparing scoring rule systems. The average difference between scores assigned to data from deceptive and truthful individuals by Backster examiners was relatively consistent and ranged from 6.24 to 7.86, while the Federal and Utah examiners scores ranged from 4.08 to 6.24 and 4.58 to 7.28, respectively. This suggests that Backster system examiners may have differed markedly from one another in absolute values assigned to data, but not in the difference between values assigned to deceptive and truthful responses. The Utah and Federal examiners in this study, on the other hand, tended to vary in conservatism when assigning values to deceptive and truthful responses. These different patterns of variability suggest that it would be easier to increase accuracy for the Utah and Federal scoring rules, which would require a consistent uniform adjustment, than for the Backster scoring rules, which would require adjustment on a per-examiner basis.

This comparison among polygraph physiological data scoring rules is arguably better than no comparison, but there are flaws. Although the data were verified regarding examinee veracity, they do not represent an exhaustive or random sample of all available data, only of the data available to us. Nor do the examiners represent an exhaustive or random sample of examiners using each system of scoring rules. Examiners were not matched for ability or experience. It would have been better to have a larger, and equal, number of examiners using all scoring rules. Ideally the examiners using the Utah scoring rules would be better trained in, and experienced at, using that system, as were the Federal and Backster examiners. The data

were collected using an Axciton polygraph, which records a hybrid measure of skin conductance (Cestaro, 1998) and does not include the photoplethysmograph channel. While adequate for the Federal and Backster examiners, the data do not include all of the information usually evaluated by examiners using the Utah system. In addition, the DoDPI ZCT is unique to Federal examiners and may have caused unusual comparisons for examiners using the Backster and Utah scoring rules. It might have also been better to ask examiners to evaluate entire question series and render veracity decisions.

In summary, the values assigned by examiners using the Backster, Federal, and Utah scoring rules were found to be equivalent in consistency and response veracity differentiation. The Utah scoring rules have the advantage of published, peer reviewed, scientific development and refinement (Raskin & Honts, 2002) which the Backster and Federal scoring rules do not have. Because of inter-scorer variability, it is likely that the Utah and Federal scoring rules could be improved using a consistent uniform adjustment, while the Backster scoring rules might require adjustment on a per-examiner basis. A much larger replication of this investigative study is necessary to determine whether the present findings are reliable. Finally, we tentatively suggest that those attempting to decide which scoring rules to use consider the practical principle of Ockham's razor—the simplest of two or more theories is preferable.

Acknowledgements

We are very grateful for the participation of the thirteen scorers in this project: Larry Barden, Karen Colonnello, Esther Harwell, Ronald Kiefer, Robert Mylott, Alvin Omohundro, Dr. Dean Pollina, Robert Randen, Terry Richie, Dr. Stuart Senter, Jimmie Swinford, Michael Woodcock, and Dr. Svetoslav Zanev. We also appreciate the data management provided by Betty Rodriguez. This effort was supported by funds from the Department of Defense Polygraph Institute as project DoDPI02-P-0005. The opinions expressed are those of the authors, and do not necessarily represent those of the U.S. Government or the Department of Defense.

References

Backster, C. (1962, Sep/Oct). Technique tips and polygraph chart interpretation. *Newsletter of the Academy for Scientific Interrogation.*

Blackwell, N.J. (1998). PolyScore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examination from actual criminal investigations (DoDPI97-R-006). Fort McClellan, AL: Department of Defense Polygraph Institute.

Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J. C. (1999). The Utah Numerical Scoring System. *Polygraph, 28*(1), 1-9.

Capps, M. H., & Ansley, N. (1992). Numerical scoring of polygraph charts: What examiners really do. *Polygraph, 21*(4), 264-320.

Cestaro, V. L. (1998). Instrumentation for presenting a known standard signal to the electrodermal activity channel for assessing response characteristics. *Polygraph, 27*(3), 188-109.

Department of Defense Polygraph Institute. (1992). *Zone comparison test.* Fort McClellan, AL: Author.

Department of Defense Polygraph Institute. (2003). *Test data analysis: DoDPI numerical evaluation scoring system* [teaching materials, For Official Use Only]. Fort Jackson, SC: Author.

Raskin, D. C., Barland, G. H., & Podlesny, J. A. (1977). Validity and reliability of detection of deception. *Polygraph, 6,* 1-39.

Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In Kleiner (Ed.) *Handbook of Polygraph Testing.* London: Academic Press.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the 7-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph, 28*(1), 10-18.

United States Army Military Police School (USAMPS), Lie Detection Committee. (1962). Chart Interpretation (DSI 406M[U]). Fort Gordon, GA: USAMPS.

Weaver, R. S. (1985). Effects of differing numerical chart evaluation systems on polygraph examination results. *Polygraph, 14*(1), 34-42.

Appendix A Participant Instructions

We thank you for your participation in our research. We believe this to be a worthwhile project, the results of which may help the profession form scoring policies. Your involvement is helping to make this possible.

The purpose of this study is to investigate the effects of certain scoring rules. In the field there are currently three principal ways in which polygraph examiners score relevant questions. One method is to score the relevant question against the stronger of the two adjacent comparison questions. This is the practice being taught at DoDPI. A second method is to always score the relevant question against the comparison question that immediately precedes it. The Utah and the Matte scoring systems take this approach. A third major method uses the Either-Or rule. The Either-Or rule is taught by Cleve Backster, and entails the decision of which comparison question to use based on whether there is a strong reaction to the relevant question. All three methods are widely used, but it is unknown whether there are differences among them. We will investigate that question with the data from you and the other volunteers.

In this project we will ask you to score 360 chart segments. All were originally recorded on Axciton polygraphs. The segments contain one relevant question bracketed by two probable-lie comparison questions. The charts are from single-issue Zone Comparison field cases of criminal suspects. The examinee was either confirmed truthful or deceptive to the relevant question you will be scoring. Please use one, and only one, of the scoring methods in the previous paragraph for all of your scoring. We also ask that you employ the 7-position scoring system.

Among these cases you will find a few of the tracings are unscoreable. These are randomly selected real world cases, and there was no attempt to include only those of textbook quality. When you find that you cannot assign a score to a tracing because of artifacts or other technical reasons, please place an "N" in the score sheet rather than a "0" or some other character. This will help us make the distinction between those reactions that were not scoreable and those where there were simply no differences between the relevant and comparison questions.

We will mail you 120 segments at a time, along with the score sheets and a copy of this introductory letter. Each packet will also include a postpaid return envelope, with which you can return the materials. We will send out a new packet as each one is mailed back to us. There is no time limit for scoring the charts. It is far better for you to give your full attention to the scoring than to rush. That said, we hope to present the data at the American Polygraph Association annual seminar next July. If your scorings are complete before June 1, 2002, they will be included in the presentation.

When we submit our manuscript for publication, you will have the option of being acknowledged by name in the report. Your decision on this matter can be deferred until the project is completed. We would also like to retain your name so that we can invite you to participate in future research.

Again, we sincerely appreciate your contribution to this project. If at any time you have any questions, please feel free to contact me at krapohld@Jackson-dpi.army.mil, or call (803) 751-5864.

DoDPI02-P-0005

Appendix B Score Sheet

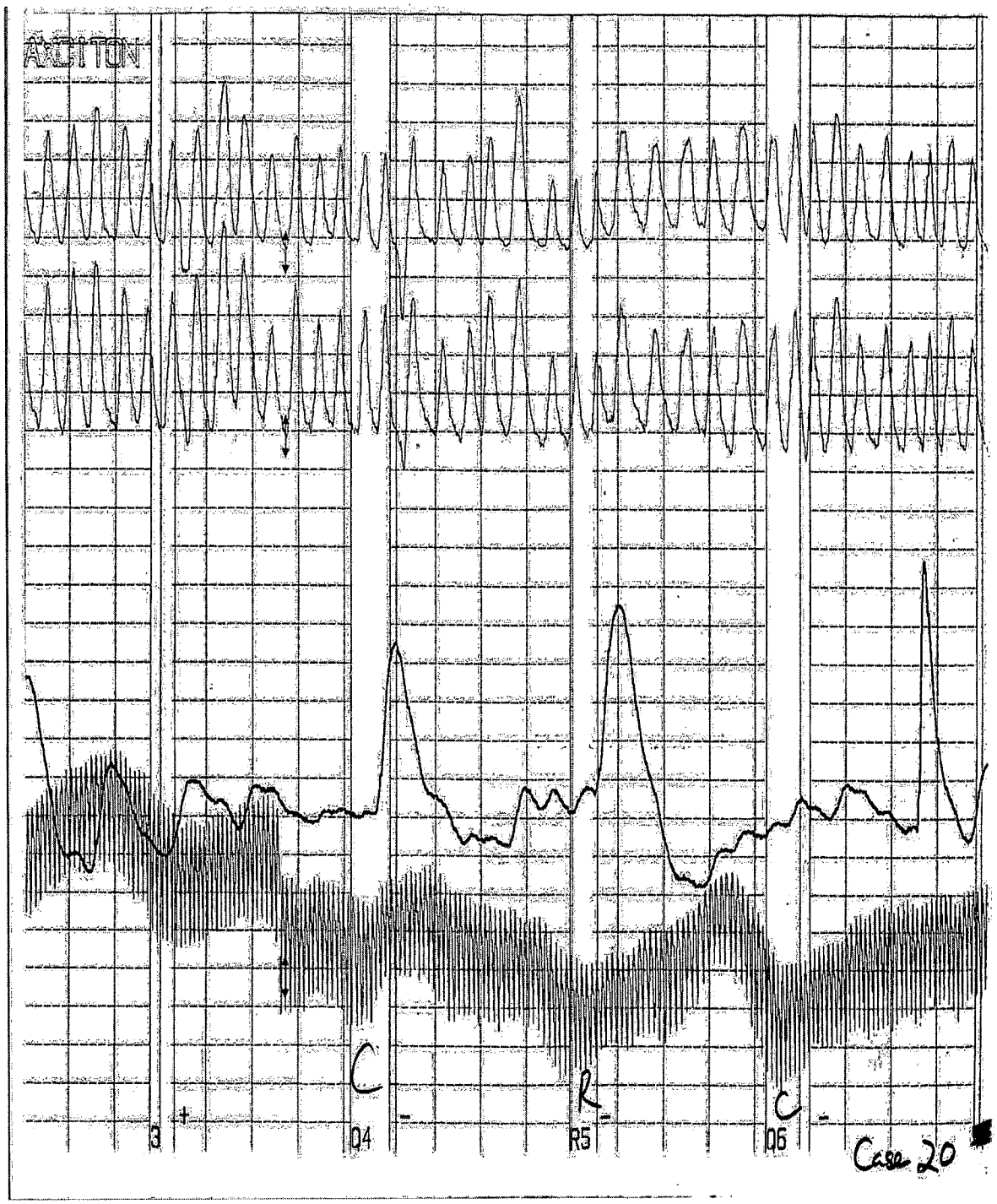
Scoring Rule Study I
DK Nov 2001

Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio
Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio
Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio
Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio
Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio
Case____	Case____	Case____	Case____
Pneumo	Pneumo	Pneumo	Pneumo
EDA	EDA	EDA	EDA
Cardio	Cardio	Cardio	Cardio

Examiner Initials______

Best Practices Project
DoDPI02-P-0005

Appendix C Chart Segment



Effects of Deception on Tonic Autonomic Arousal

John C. Kircher, Ted Packard, Brian G. Bell, & Paul C. Bernhardt

Abstract

The present study tested if measures of tonic arousal are related to the amplitude of responses during probable-lie and directed lie polygraph examinations. It also tested if tonic levels of electrodermal and cardiovascular activity can be used to improve the accuracy of polygraph examinations. Three hundred and thirty-six male and female participants in a previous experiment (DODPI97-P-0016) were interrogated about their participation in a mock crime. Half of the subjects were guilty of committing the mock crime and half were innocent. Half of the innocent subjects and half of the guilty subjects received a probable-lie polygraph examination. The remaining subjects received a directed lie polygraph examination. Subjects were offered a $50 bonus to appear truthful on the polygraph examination. Tonic and phasic measures of skin conductance and skin resistance were obtained from skin conductance recordings. Tonic and phasic systolic and diastolic blood pressure measures were obtained from a Finapres® blood pressure monitor. Tonic and phasic measures of heart period were obtained from the electrocardiogram (EKG). Consistent with prior research, two tonic measures of skin conductance were positively related to phasic reactivity. Weak but significant correlations between tonic and phasic arousal also were obtained for blood pressure and heart period. However, none of the tonic measures improved the accuracy of polygraph outcomes. Tonic arousal accounted for less than 2% of the variance in the guilt/innocence criterion when used in combination with standard measures of differential reactivity to predict group membership. The results suggest that the use of absolute measures of electrodermal and cardiovascular activity would do little to improve the accuracy of computer algorithms for diagnosing truth and deception.

Introduction

The primary objective of the present study was to determine if tonic levels of skin conductance and other measures of tonic arousal can be used to improve the accuracy of probable-lie and directed lie polygraph tests. The present study also assessed the extent to which tonic skin conductance and other measures of tonic arousal (skin resistance, arterial blood pressure, and heart period) are related to the magnitude of responses during probable-lie and directed lie polygraph examinations.

Background

It is well known that there are large differences among individuals in both tonic levels of physiological arousal and in the magnitude of phasic responses to stimuli. Some polygraph subjects will have basal skin conductance levels that measure less than one μSiemen (1 M ohm), whereas other subjects in the same situation will have skin conductance levels that approach 100 μSiemens (10 K ohm; Venables & Christie, 1980). Nonspecific skin conductance responses, another measure of tonic arousal, may range from zero to 10 per minute (Boucsein, 1992). Tonic heart rate ranges from 50 BPM to over 120 BPM, and tonic levels of mean arterial blood pressure range from 70 mm Hg to over 130 mm Hg (Rushmer, 1976).

Computer and numerical scoring procedures remove individual differences in tonic arousal by making within-subject comparisons of the individual's physiological reactions to different types of test questions. If a reaction is noticeably greater to one question than to another, the larger reaction is considered diagnostic (Raskin, 1989). To make these judgments, the computer or polygraph examiner considers only the relative strength of reactions to different types of test questions. No systematic attempt is made to account for

the fact that reactions to test questions are superimposed on a baseline of tonic activity.

Psychophysiological research indicates that tonic levels of activation correlate with the magnitude of evoked responses to stimuli (e.g., Hord, Johnson, & Lubin, 1964). Tonic arousal also predicts the habituation of responses that typically occurs with repeated presentations of a stimulus. Katkin (1975) found that subjects with high levels of electrodermal activity showed less habituation to a series of tones than did subjects with low levels of tonic electrodermal activity.

However, results from studies are mixed in which tonic levels were related to differential responses to signal and nonsignal stimuli. Some studies suggest that tonic levels predict differential responses to signal and nonsignal stimuli (Katkin, 1975), whereas others do not (Schell, Dawson, & Filion, 1988). The relationship between tonic arousal and differential reactivity to signal and nonsignal stimuli is important because polygraph examiners base their decisions on differences between reactions to test questions that differ in signal value (Raskin, 1979). The relationship between tonic arousal and habituation is important because a polygraph examiner may present the same basic question as many as 15 times over the course of a polygraph examination, and the subject's physiological responses tend to habituate. If measures of tonic arousal predict individuals' patterns of response during a polygraph examination, they might be used as statistical 'filters' to remove noise from physiological measures that are used to predict the criterion. In so doing, tonic arousal may enhance the ability of those measures to discriminate between truthful and deceptive subjects.

The possibility that individual differences in tonic arousal can be used to improve the accuracy of polygraph tests has never been investigated. The issue remains unexplored primarily because, with the exception of heart rate, traditional polygraph instrumentation provides only relative measures of change in physiological activity. Absolute measures of physiological activity are needed to determine if individual differences in tonic arousal can be used to increase decision accuracy.

The CPS-LAB system used to collect the polygraph charts for the present study allowed us to measure skin conductance in microSiemens (μS), skin resistance in ohms, heart period in ms, and blood pressure in mm Hg. The availability of absolute measures of tonic activity allowed us to investigate the possibility that this large untapped resource of physiological information might be used to increase the accuracy of probable-lie and directed lie tests.

The goals of the present study were twofold. First, based on research by Hord et al. (1964), we predicted that tonic levels of autonomic arousal would be related to the strength of phasic reactions to test questions observed during probable-lie and directed lie examinations. The second objective was to determine if tonic autonomic arousal could be used to improve the accuracy of probable-lie and directed lie tests.

Methods

Subjects

Four-hundred and seventeen adults were recruited from the general community by newspaper advertisements for a study that examined the effects of the demonstration test on the accuracy of probable-lie and directed lie polygraph examinations (DoDPI97-P-0016). The advertisements offered $30 for two hours of participation and the opportunity to earn an additional $50 bonus. Of the 417 individuals, 81 were eliminated from the study for a variety of reasons. Thirty-three subjects assigned to the guilty condition (16%) declined to participate after they received their instructions to commit a simulated theft. Eighteen individuals failed to follow instructions (e.g., did not commit the theft yet reported for their polygraph, arrived late, brought a child with them to the lab). Thirteen individuals were dismissed due to health problems. Health problems included reports of pain, less than four hours of sleep, and high blood pressure. Nine individuals assigned to the guilty condition (5%) confessed. Equipment problems and experimenter errors resulted in the loss of eight other individuals. The remaining 168

innocent and 168 guilty subjects were retained to fill the cells of the design matrix (described below).

The mean age of the sample was 30.7 years (*SD* = 11). Years of education ranged from 9 to 25 (*M* = 14.3, *SD* = 2.5). Most participants were Caucasian (87.5%), 5.7% were Hispanic, and the remaining 6.8% were Black, Asian, American Indian, or chose not to respond. Fifty-three percent of the participants were single, 33.9% were married, and the remaining 12.2% were divorced, separated, or widowed. Although a wide range of occupations was represented, over 75% of the sample fell into one of the following eight categories: student (17%), professional (11.9%), sales worker (9.2%), office worker (8.3%), service worker (8.3%), unemployed (7.7%), homemaker (7.7%), or laborer (7.4%).

Design

Guilty and innocent subjects were randomly assigned to one of 16 cells in a completely crossed 2 x 2 x 4 factorial design with equal numbers of male and female participants in each cell. The design is illustrated in Figure 1. All factors except Sex are represented in the figure.

The first factor, Guilt, had two levels; 168 participants were guilty of committing a mock crime and the remaining 168 were innocent of the crime. The second factor, Test Type, also had two levels; half of the participants were given probable-lie comparison question tests (PL) and half were given directed lie tests (DL).

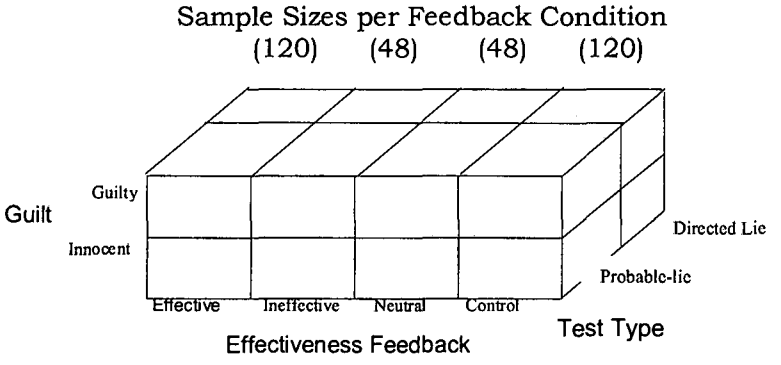


Figure 1. Design of experiment

The third factor, Effectiveness Feedback, had four levels. Participants were unevenly distributed over the four levels of the Effectiveness Feedback factor. One group of 120 participants (30 participants in each of the four cells shown on the far left of Figure 1) received the type of feedback commonly provided to subjects in actual field examinations. Prior to their polygraph test, they were given a demonstration test and told, regardless of the outcome, that they showed their strongest reaction to the number they had chosen. They also were told they should have no problem passing the polygraph test if they were completely truthful to all of the questions (Effective Feedback group).

Twelve participants were assigned to each of the four Ineffective Feedback cells of the design matrix. Participants who received ineffective feedback were given a numbers test and were told, regardless of the outcome, that they did not react appropriately to the chosen number. They also were told that it would be difficult to determine if they were lying or telling the truth during their polygraph test.

Twelve participants were assigned to each of the four Neutral Feedback cells of the matrix. Participants who received neutral feedback were given a numbers test and were told that the test would provide an opportunity for the participant to practice answering

questions and for the polygraph examiner to adjust the instrument. Participants were given no information about the outcome of the numbers test.

Thirty participants were assigned to each of the four control groups illustrated on the far right of Figure 1. The pretest procedures for subjects in the control groups were the same as those used for other subjects except that control subjects were not given a numbers test.

To summarize, 120 participants were given the demonstration test and received feedback that the test was effective. Another 48 participants were given a demonstration test and received feedback that the test was ineffective. Another 48 participants were given a demonstration test and received neutral feedback. The remaining 120 participants were not given a numbers test. Within each level of the Feedback factor, the design was balanced in terms of numbers of guilty and innocent male and female subjects who were given either probable-lie or directed lie polygraph examinations.

Two examiners administered all of the polygraph examinations. One examiner was an advanced doctoral student in educational psychology. The graduate student (PCB) tested 12 subjects in each of the 16 cells in the design matrix (192 subjects). The post-doctoral research associate (BGB) tested the remaining 144 subjects. The principal investigator (PI) trained both of the examiners. The PI has been conducting research on polygraph techniques at the University of Utah for the past 24 years and participated in annual workshops to train professional polygraph examiners for 17 of those years.

Procedures

The procedures followed those described elsewhere (Kircher & Raskin, 1988). Prospective participants called a secretary who screened the participants for eligibility and briefly described the experiment and pay policy. Callers were invited to participate if they met the following criteria: (1) they were between 18 and 65, (2) they were not taking prescription medication, (3) they had never had a polygraph test, (4) they were fluent in English, and (5) they had no major medical problems.

Callers who agreed to participate were given an appointment to report to a room in a building on the campus of the University of Utah. When the participant arrived, an envelope addressed to the participant was taped to the door. Instructions within the envelope directed the participant to enter the room, close the door, read and sign an informed consent form, complete a brief questionnaire, and then play a cassette recorder that presented their instructions over headphones.

Guilty participants received tape-recorded instructions to commit a mock theft of a $20 bill from a wallet that was in a purse in a desk in a secretary's office. Participants went to a secretary's office on a different floor of the building where they asked the secretary where Dr. Mitchell's office was located. The secretary was actually a confederate in the experiment. The secretary responded that there was no Dr. Mitchell in the department. The participant thanked the secretary and left the office. The participant then waited in the hallway until the secretary left the office unattended (1-3 minutes), entered the office, searched the desk for the purse, and took the $20 bill from the wallet that was in the purse. Participants were instructed to conceal the $20 on their person and go to a room where they waited for the polygraph examiner. Guilty participants were instructed to prepare an alibi in case they were caught in the office. Innocent participants listened to a general description of the crime, left the area for 15 minutes, and went to a room where they waited for the polygraph examiner.

All participants were told that they would be given a polygraph test by a polygraph expert who did not know whether they had committed the theft. They were told that the examiner would use a computer to assist in the analysis of their polygraph charts, and if they could convince the polygraph examiner of their innocence, they would receive $80. They were also told that if they failed to convince the examiner of their innocence, they would only receive $30.

After the participant had reported to the waiting room, the polygraph examiner went to the room, introduced himself, and instructed the participant to go to the restroom and wash their hands with soap and warm water. The participant was then escorted to the lab where the examiner obtained some biographical information and attached the sensors to the participant. Standard field polygraph procedures were used, and the session was videotaped and audiotaped. The polygraph examiner described the role of the autonomic nervous system in the detection of deception. He then described and administered the numbers test to participants in the effective, ineffective, and neutral feedback conditions. Finally, the polygraph examiner reviewed the appropriate set of test questions with the participant. Relevant questions that pertained to the theft and the sacrifice relevant question were reviewed first, probable-lie or directed lie comparison questions were reviewed next, and the neutral and outside issue questions were reviewed last. The test questions for participants assigned to the probable-lie condition were as follows:

(Outside Issue) 1. Do you understand that I will ask only the questions that we have discussed?

(Sacrifice Relevant) 2. Do you intend to answer truthfully all of the questions regarding the theft of the $20?

(Neutral) 3. Do you live in the United States?

(Probable-lie) 4. Before the age of ___, did you ever take something that didn't belong to you?

(Relevant) 5. Did you take that $20 from the purse?

(Neutral) 6. Is today ____?

(Probable-lie) 7. During the first ____ years of your life, did you ever do anything that was dishonest or illegal?

(Relevant) 8. Did you take that $20?

(Neutral) 9. Is your first name ___?

(Probable-lie) 10. Between the ages of ___ and ___, did you ever lie to get out of trouble?

(Relevant) 11. Do you have that $20 with you now?

The age used for the probable-lie questions was one year less than the subject's current age to exclude the issue under investigation from consideration.

The test questions for participants assigned to the directed lie condition were the same as those presented to participants in the probable-lie condition, except that the probable-lie questions in positions 4, 7, and 10 were replaced with the following directed lie questions:

(Directed Lie) 4. In your entire life, did you ever tell even one lie?

(Directed Lie) 7. Have you ever broken a rule or regulation?

(Directed Lie) 10. Did you ever make a mistake?

After reviewing the test questions, a probable-lie or directed lie test was administered. The interval between question onsets was at least 25 s, and the interval between repetitions of the question sequence was between one and three minutes. For probable-lie participants, after the first chart, the examiner asked the participant if they noticed anything unusual during the polygraph examination. After the second chart, the examiner asked the participant if they noticed anything unusual when they were asked one of the probable-lie questions. For directed lie participants, after each of the first three charts the examiner asked the participant if they were lying to the directed lie items and if they felt any differently when they lied. These procedures were designed to draw the participant's attention to the comparison questions, and reduce the risk of false positive errors.

The question sequence was presented five times. Neutral and comparison questions were rotated over repeated presentations of the question sequence such that each relevant question was preceded by each neutral and

each comparison question at least once. The orders of presentation of the questions were not reviewed with the participant in advance.

At the conclusion of the test, the sensors were removed, and the subject was asked to complete posttest questionnaires. The probability that the participant was truthful was then computed from the physiological responses using algorithms described elsewhere (Kircher & Raskin, 1988). If the probability of truthfulness exceeded 0.70, the participant was paid $80, $30 for their time and a $50 bonus. Otherwise, the participant was paid $30. The participant was then debriefed and released.

Apparatus

The CPS-LAB system (Scientific Assessment Technologies, SLC, UT) was used to configure the data collection hardware, specify storage rates for the physiological signals, and build automated data collection protocols. CPS-LAB also was used to collect, edit, and score the physiological data.

The physiological data acquisition subsystem (PDAS) of CPS-LAB generated analog signals for thoracic and abdominal respiration, skin conductance, cardiograph, finger pulse amplitude, skin potential, and cardiotachometer. In addition, calibrated analog output from a Ohmeda 2300 Blood Pressure Monitor was routed to a general-purpose coupler on the PDAS. Each of the eight analog signals was digitized at 1000 Hz with a Metrabyte DAS 16F analog-to-digital converter installed in a 50 MHz PC compatible 486 computer with 16 MB of RAM.

Respiration was recorded from two Hg strain gauges secured with Velcro straps around the upper chest and the abdomen just below the rib cage. The strain gauge changed in resistance as the subject breathed. Resistance changes were recorded DC-coupled with a 2-pole, low-pass filter, f_c = 13 Hz.

Skin conductance was obtained by applying a constant voltage of .5V to two UFI 10mm Ag-AgCl electrodes filled with .05M NaCl in a Unibase medium. The electrodes were taped with adhesive collars to the distal phalanx of the ring and last fingers of the left hand. The signal was recorded DC-coupled with a 2-pole, low-pass filter, f_c = 6 Hz.

The cardiograph was recorded from a blood pressure cuff wrapped around the right upper arm and inflated to 55 to 60 mm of Hg at the beginning of each chart. The cuff was connected by rubber tubing to a Motorola MPX10DP pressure transducer in the PDAS. The output from the pressure transducer was amplified and recorded DC-coupled with a 2-pole, low-pass filter, f_c = 8.8 Hz.

Finger pulse amplitude was obtained from a UFI photoplethysmograph attached to the first finger of the left hand with a Velcro strap. The signal from the photocell was AC-coupled with a 0.2-second time constant and a 2-pole, low-pass filter, f_c = 10 Hz.

The electrocardiogram was obtained from Lead II using disposable, pre-gelled Red Dot™ Ag-AgCl snap electrodes taped to the left arm and right leg. The PDAS generated a 20 ms square wave pulse that coincided with the R-wave in the electrocardiogram. The square wave from the PDAS was routed to the analog-to-digital converter, and the CPS-LAB software measured and stored the time between successive pulses (interbeat intervals).

Skin potential was recorded from Beckman 10mm Ag-AgCl electrodes filled with .05 M NaCl in a Unibase medium attached to the thumb of the left hand (active site) and the lower arm, just below the elbow (inactive site). The inactive site was rubbed with alcohol prior to applying the electrode. Skin potential was recorded DC-coupled with a 2-pole, low-pass filter, f_c = 10 Hz. A 39.2 K ohm resistor was soldered in series with the reference (inactive) electrode to prevent variations in skin potential from affecting the skin conductance recordings.

The finger cuff of the Finapres Blood Pressure Monitor was attached with Velcro to the middle phalanx of the middle finger on the left hand. Continuous calibrated voltage changes from the Finapres Monitor were routed to a general purpose coupler on the PDAS where it was recorded DC-coupled with a 2-pole, low-pass filter, f_c = 10 Hz. The voltage changes were converted to absolute blood pressure in mm of Hg.

The 1000 Hz samples for each channel were reduced prior to storing them on the hard disk by averaging the samples for successive epochs. Respiration and electrodermal channels were stored at 10 Hz. Cardiograph, finger pulse, and blood pressure signals were stored at 100 Hz. The cardiotachometer produced an interbeat interval measured to the nearest ms for each heart beat.

Calibration Procedures

To assess the relationships between tonic arousal and phasic reactivity to test questions, it was necessary to convert the raw data in analog-to-digital converter units to absolute units for skin conductance, skin resistance, and blood pressure. The CPS-LAB system already provided the interbeat intervals in ms that were required to study heart period and vagal tone. For skin conductance, a separate multiple regression equation was developed for each of six possible gain settings on the PDAS. Each equation predicted known conductances from the offset on the front panel, internal PDAS digital-to-analog (DAC) offset settings, and observed analog-to-digital converter values. The conductance values used to calibrate the instrument ranged from 1 μSiemen (1 M ohm) to 50 μSiemens (20 K ohm). External (front panel) and internal (DAC) offsets were also systematically varied to ensure that the resulting equation would work for any configuration of gain and offset settings. Each equation accounted for over 99.8% of the variance in known inputs.

Since resistance (R) is the reciprocal of conductance (G), skin resistance was obtained by inverting the calibrated skin conductance signal prior to extracting measurements of response amplitude, i.e., $R = 1/G$.

The methods used to derive absolute measures of skin conductance were used to develop equations to measure the output voltages generated by the Finapres. The Finapres generated a voltage that ranged from 0V to 2V that was linearly related to blood pressure that ranged from 0 mm Hg to 200 mm Hg. Again, the error of measurement was negligible; the regression equations accounted for over 99.8% of the variance in the voltages obtained from the Finapres.

Measurements of Autonomic Activity

Tonic Arousal

For each autonomic measure, tonic levels were measured by calculating the mean of 5-second epochs of basal activity prior to the onset of each question within each chart. The frequency of nonspecific skin conductance responses was measured in addition to measuring skin conductance level. Nonspecific responses were measured during the last 15 seconds of the recording interval that followed each neutral question.

Preliminary examination of the distributions of all tonic measures of arousal revealed, as expected, significant positive skew for measures of skin conductance level and number of skin conductance responses. Log transformations of the skin conductance measures normalized the distributions (Venables & Christie, 1980). To avoid undefined values (log of zero), a value of 1 was added to each skin conductance score prior to taking the log of the score.

Phasic Reactivity

Measures of phasic reactivity included peak amplitude of the skin conductance response (SCR), peak amplitude of the skin resistance response (SRR), peak amplitude of increases in systolic blood pressure (SBPR), and peak amplitude of increases in diastolic blood pressure (DBPR). For heart period, the mean heart period for the last 2 beats prior to question onset was subtracted from the longest heart period between 4 and 15 sec after question onset (HPR, Podlesny & Kircher, 1999). A log transformation of SCR also was performed to normalize its distribution.

Indices of Differential Reactivity to Comparison and Relevant Questions

Following our standard protocol (Kircher & Raskin, 1988), an index of differential reactivity to comparison and relevant questions was computed for each subject and each autonomic measure. For example, each subject provided 18 measurements of skin conductance amplitude for the three comparison questions and the three relevant questions on each of the first three charts. The 18 measurements were converted to Z scores. The mean of the nine Z

scores for relevant questions was then subtracted from the mean of the nine Z scores for comparison questions.

Indices of differential reactivity can be weighed and combined by means of a discriminant function or regression equation to maximize discrimination between truthful and deceptive subjects (Kircher & Raskin, 1988; Kircher, Woltz, Bell, & Bernhardt, 1998). An index of differential reactivity is analogous to the total numerical score assigned by the polygraph examiner for a particular channel. The index was positive when the mean reaction to comparison questions was greater than the mean reaction to relevant questions, and the index was negative when the reactions to relevant questions were greater. Since innocent subjects were expected to react more strongly to comparison questions and guilty subjects were expected to react more strongly to relevant questions, we expected positive scores for innocent subjects and negative scores for guilty subjects.

For all variables except respiration, a large measured response was indicative of a strong reaction. For respiration excursion, suppressed respiratory activity was indicative of a strong reaction. Thus, innocent subjects were expected to show relatively small measured respiration responses (suppression) to comparison questions, whereas guilty subjects were expected to show relatively small measured respiration responses (suppression) to relevant questions. To maintain consistency of interpretation across physiological measures, the sign of the index of differential reactivity for respiration was reversed.

Results

Preliminary Analyses

Treatment-Related Attrition

Thirty-three individuals assigned to the guilty condition (16%) refused to participate after they had received their tape-recorded instructions, whereas none of the innocent subjects declined to participate. Consequently, subjects who agreed to commit the mock crime may have been sampled from a population that differed in certain respects from the more general population from which innocent subjects were drawn. For example, subjects who remained in the guilty condition on average may have been older or less anxious than subjects in the innocent condition. Preliminary tests were conducted to explore the possibility that guilty and innocent groups differed on measures of marital status, ethnicity, occupation, age, education, hours of sleep, the Marlowe-Crowne scale (Crowne & Marlowe, 1964), Rotter Trust scale (Rotter, 1967), and two anxiety scales (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The guilty and innocent subjects who completed the experiment did not differ significantly on any of the demographic or personality measures.

Effects of Feedback

The design of the original study (DoDPI97-P-0016) included feedback conditions that were not representative of current field practice. Factorial ANOVAs were conducted to determine if the effects of interest varied as a function of Feedback and other facets of the design. The results of the ANOVAs and effect sizes are summarized in Table 1.

Since Question Type (comparison versus relevant) was not a factor in these analyses, no significant effects were expected, with one possible exception. The effect of Guilt obtained for heart period level (HPL) is consistent with data from a field study reported by Krapohl and Ansley (1999). Guilty subjects had significantly shorter heart periods (higher heart rates) than did innocent subjects. In general, however, no more than 4% of the variance in any tonic or phasic measure was related to Feedback or any interaction between Feedback and Guilt, Test Type, or Sex. In addition, the number of significant effects was no greater than what one would expect to occur by chance.

Since an objective of the present research was to assess the *relationship* between tonic arousal and phasic reactivity, preliminary tests also were conducted for heterogeneity of regression slopes among the various treatment conditions. If the relationship between tonic arousal and reactivity varied as a function of Feedback, then the research plan was to analyze data only from the effective feedback groups. The plan was to focus on the effective feedback

groups because the procedures used for those groups were most similar to the procedures used in actual field polygraph examinations.

Separate tests for heterogeneity of regression were conducted for skin conductance, skin resistance, systolic and diastolic blood pressure, and heart period. The results revealed no evidence that the slopes of the regression lines that related phasic reactivity to tonic level varied across the 16 cells in the design matrix (see Figure 1). Nor was there evidence that the relationships between differential reactivity and tonic arousal varied across treatment conditions.

Table 1. *Proportions of Variance (η^2) in Tonic and Phasic Measures Explained by Guilt, Test Type, Feedback, and Sex*

Factor	SCN[1]	SCL[2]	SCR[3]	SRL[4]	SRR[5]	SBPL[6]	SBPR[7]	DBPL[8]	DBPR[9]	HPL[10]	HPR[11]
G-uilt	.00	.01	.01	.01	.00	.00	.02*	.00	.02*	.03*	.01
T-est	.01	.02*	.01	.01	.05*	.00	.01	.00	.02*	.00	.00
F-eedback	.03*	.01	.00	.01	.00	.00	.02	.00	.02	.00	.00
S-ex	.00	.02*	.00	.02*	.03*	.00	.00	.04*	.01	.01	.00
GxT	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
GxF	.00	.01	.00	.00	.01	.01	.00	.01	.02	.01	.01
GxS	.00	.00	.00	.00	.00	.00	.02*	.00	.02*	.00	.01
TxF	.01	.01	.01	.01	.02	.02	.02	.02	.04*	.00	.01
TxS	.00	.00	.00	.00	.01	.00	.00	.00	.01	.00	.00
FxS	.00	.03*	.02	.01	.01	.02	.03*	.03*	.00	.01	.01
GxTxF	.00	.01	.01	.02	.01	.02	.01	.02	.02	.02	.02
GxTxS	.01	.01	.00	.00	.00	.01	.01*	.00	.00	.00	.00
GxFxS	.01	.01	.01	.01	.00	.00	.01	.00	.01	.00	.01
TxFxS	.02	.01	.01	.01	.00	.00	.01	.00	.00	.04*	.02
GxTxFxS	.01	.00	.00	.00	.00	.00	.02	.01	.02	.01	.00

[1] Log Number of Nonspecific Skin Conductance Responses
[2] Log Skin Conductance Level
[3] Log Skin Conductance Response
[4] Skin Resistance Level
[5] Skin Resistance Response
[6] Systolic Blood Pressure Level
[7] Systolic Blood Pressure Response
[8] Diastolic Blood Pressure Level
[9] Diastolic Blood Pressure Response
[10] Heart Period Level
[11] Heart Period Response
* $p < .05$

In summary, the results shown in Table 1 revealed little or no difference among the groups on mean levels of tonic arousal and reactivity. The tests for heterogeneity of regression also revealed no reliable differences among the groups in the relationships between arousal and reactivity. The results of these preliminary analyses suggested that the relationships of interest were the same or similar across the various facets of the design. For example, the relationship between tonic arousal and phasic reactivity for guilty subjects in the effective feedback condition did not differ significantly from the relationship between tonic arousal and phasic reactivity for innocent subjects who received neutral or ineffective feedback.

To maximize power and precision, subsequent analyses were conducted with the entire sample of 336 cases. However, partially redundant, parallel tests were conducted using only the subjects in the effective feedback conditions. These tests were conducted to verify that the effects observed in the entire sample were consistent with those

from the subjects who were given the numbers test and received effective feedback (n = 120).

Table 2. *Means (and Standard Deviations) of Physiological Measures*

Physiological Measure	*Entire Sample (N=336)*		*Probable-lie Effective Feedback (n=60)*		*Directed Lie Effective Feedback (n=60)*	
Skin Conductance Log Number of Responses	.235	(.187)	.203	(.185)	.187	(.173)
Skin Conductance Log Level (μS)	.611	(.238)	.634	(.230)	.557	(.240)
Skin Conductance Log Amplitude (μS)	.144	(.119)	.118	(.106)	.157	(.148)
Skin Conductance Differential Reactivity	-.034	(.792)	-.201	(.876)	-.115	(.825)
Skin Resistance Tonic Level (K ohms)	291.36	(190.24)	277.26	(163.21)	327.01	(195.63)
Skin Resistance Amplitude (K ohms)[b]	-19.65	(19.10)	-13.51	(11.26)	-26.65	(28.77)
Skin Resistance Differential Reactivity	-.041	(.839)	-.236	(.907)	-.131	(.881)
Systolic Blood Pressure Tonic Level (mm of Hg)[a]	103.69	(15.49)	100.37	(17.53)	106.14	(12.86)
Systolic Blood Pressure Amplitude (mm of Hg)[a]	13.56	(4.98)	12.44	(4.38)	14.18	(4.87)
Systolic Blood Pressure Differential Reactivity	.013	(.686)	-.067	(.657)	.012	(.662)
Diastolic Blood Pressure Tonic Level (mm of Hg)	62.7	(10.91)	61.89	(12.85)	64.03	(9.90)
Diastolic Blood Pressure Amplitude (mm of Hg)[b]	7.49	(2.36)	6.66	(1.70)	7.93	(2.20)
Diastolic Blood Pressure Differential Reactivity	.055	(.679)	-.018	(.693)	-.011	(.684)
Heart Period Tonic Level (msec)	760.22	(120.71)	777.95	(120.92)	758.03	(130.84)
Heart Period Phasic Response (msec)	50.18	(33.44)	53.12	(32.98)	50.23	(34.01)
Heart Period Differential Reactivity	-.257	(.661)	-.071	(.718)	-.241	(.644)

[a]The difference between the means for probable-lie and directed lie tests was significant at $p < .05$.
[b]The difference between the means for probable-lie and directed lie tests was significant at $p < .01$.

Table 2 shows the mean and standard deviation of each physiological measure for the entire sample of 366 subjects and for probable-lie and directed lie subjects who received effective feedback. Subjects in the probable-lie condition showed larger skin

resistance responses, whereas subjects in the directed lie condition had higher systolic blood pressure levels and produced stronger systolic and diastolic blood pressure responses.

Correlations Between Tonic Arousal and Phasic Reactivity

One objective of the present research was to determine if tonic levels of autonomic arousal are related to the magnitude of physiological reactions to test questions. Table 3 shows the correlation between each tonic measure of autonomic activity and the corresponding mean amplitude of the phasic response for all cases combined and for the two effective feedback conditions.

Table 3. *Correlations Between Measures of Tonic Level and Phasic Reactivity*

Tonic Arousal	*Entire Sample (N=336)*	*Probable-lie Effective Feedback (n=60)*	*Directed Lie Effective Feedback (n=60)*
Skin Conductance Log Nonspecific Responses	.661 **	.758 **	.717 **
Skin Conductance Log Level	.580 **	.596 **	.610 **
Skin Resistance Level	-.199 **	-.079	-.227
Blood Pressure Systolic Level	.102	.189	.181
Blood Pressure Diastolic Level	.114 *	-.077	.128
Heart Period Level	.138 *	.133	.052

* $p < .05$
** $p < .01$

As shown in Table 3, the strongest correlations were between measures of tonic and phasic electrodermal activity. For skin conductance, high tonic levels were associated with relatively strong skin conductance responses. The correlation between the log number of nonspecific responses and the log skin conductance level was .57 for the entire sample. Since skin resistance is the inverse of skin conductance, the negative correlation for skin resistance was expected. High tonic levels of skin resistance were associated with relatively small skin resistance responses.

For the entire sample, a small but significant positive correlation also was observed for heart period. Long interbeat intervals were associated with large increases in heart period. Stated differently, low basal heart rates were associated with the greatest drops in heart rate following question onset. High blood pressure levels tended to be positively related to the magnitude of the phasic response, but the correlations were generally small and may not be reliable.

Correlations Between Tonic Arousal and the Criterion (Guilt)

Another objective of the present study was to determine if measures of tonic arousal could be used to improve the accuracy of probable or directed lie polygraph tests. The

first method used to assess the potential usefulness of each tonic arousal measure was to calculate a point-biserial correlation between the measure and the criterion. The criterion was a dichotomous variable that distinguished between guilty (coded 0) and innocent subjects (coded 1; Kircher & Raskin, 1988). Table 4 shows these point-biserial correlations for the entire sample and the effective feedback conditions separately.

Table 4. *Point-biserial Correlations Between Measures of Tonic Level and the Guilt/Innocence Criterion*

Physiological Measure	*Entire Sample (N=336)*	*Probable-lie Effective Feedback (n=60)*	*Directed Lie Effective Feedback (n=60)*
Skin Conductance Log Number of Responses	-.042	.047	.011
Skin Conductance Log Level	-.093	.013	-.089
Skin Resistance Level	.110 *	.057	.136
Systolic Blood Pressure Level	-.106	-.271 *	.045
Diastolic Blood Pressure Level[a]	-.031	-.142	.251
Heart Period Level	.206 **	.424 **	.092

* $p < .05$
** $p < .01$
[a]The difference between the correlations for probable-lie and directed lie effective feedback conditions was significant at $p < .05$.

Most of the correlations with the criterion were not significant. The greatest correlations were obtained from subjects who were given probable-lie tests. In that condition, guilty subjects had higher heart rates (M = 84.3 BPM) than did innocent subjects (M = 73.8 BPM; r = .42). Guilty probable-lie subjects also had higher systolic blood pressure levels (M = 105.1) than did the innocent subjects (M = 95.7; r = -.27). There were no significant differences in tonic arousal between guilty and innocent subjects who received directed lie tests.

Regression Equations that Combine Tonic Level and Standard Measures to Predict the Criterion

Several types of multiple regression analyses were conducted to determine if measures of tonic arousal could be used in combination with traditional indices of differential reactivity to improve discrimination between truthful and deceptive individuals. For each type of analysis, a tonic measure of arousal was added to a base model, and the contribution of the tonic measure to the prediction equation was tested for significance. In all of these analyses, the criterion to be predicted was a dichotomous variable that

distinguished between guilty (coded 0) and innocent subjects (coded 1).

The predictor variables in the base model were indices of differential reactivity. The base model contained one index of differential reactivity for each response system. Each base model contained a respiration index, an electrodermal index, and a cardiovascular index. The particular set of base model predictors depended on the tonic measure of arousal that was to be tested for significance. The base model contained an index of differential reactivity from the same channel as the one that provided the measure of tonic arousal. For example, when the tonic measure of arousal was systolic blood pressure level, the index of differential reactivity for systolic blood pressure responses was substituted for the cardiograph in the base model. We used this approach because we expected that tonic systolic blood pressure would correlate more highly with its own corresponding index of differential reactivity than with some other measure of cardiovascular reactivity.

The stronger the correlation between measures of tonic arousal and differential reactivity, the more likely it was that the tonic measure would serve as a suppressor variable and would make a significant contribution to the prediction equation. Suppression occurs when one variable in the model (tonic level) is uncorrelated with the criterion, but it is highly correlated with another variable (differential reactivity index) that *is* correlated with the criterion (Cohen & Cohen, 1975). A suppressor variable improves the diagnostic validity of a predictor variable by removing noise from the predictor variable that attenuates the predictor variable's correlation with the criterion. In general, the best suppressor variable is one that correlates near zero with the criterion, yet it is highly correlated with another variable in the regression equation.

Linear Analysis of Tonic Measures of Arousal

Table 5 summarizes the results from the first set of regression analyses. The first column of Table 5 lists the indices of differential reactivity included in the base model. The second column shows the tonic measure of arousal that was added to the base model. The third column shows the proportion of variance in the criterion explained by the base model for the entire sample of 336 subjects (R^2). The fourth column shows the increment in the proportion of variance explained by the measure of tonic arousal (ΔR^2). The last four columns show the R^2 and the ΔR^2 for probable-lie and directed lie subjects who received effective feedback.

The results in Table 5 indicate that the various measures of tonic arousal added little to the regression equations. The only measure of tonic arousal to make a significant contribution to the base model was heart period level. Discriminant analyses were performed to compare the accuracy of dichotomous classifications into truthful and deceptive groups for the base model and the model that included heart period level. Although the contribution of heart period level to the regression equation was statistically significant, the classification accuracy was slightly lower for the model that included heart period level (80.1%) than for the base model (81.5%).

Nonlinear Analysis of Tonic Measures of Arousal

Indices of differential reactivity reflect the extent to which subjects responded more strongly to comparison or relevant questions. Although tonic measures of electrodermal and cardiovascular activity were linearly related to the magnitude of phasic reactions to test questions (see Table 3), tonic measures may be nonlinearly related to the *differences* in reactions to comparison and relevant questions. Subjects who react strongly to comparison questions and have large positive indices of differential reactivity may have relatively high tonic levels of activity. Similarly, subjects who react strongly to relevant questions and have large negative indices of differential reactivity also may have high tonic levels of activity. Conversely, subjects who show little difference in their reactions to comparison and relevant questions may have relatively low levels of tonic activity. Under these conditions, a nonlinear transformation of tonic arousal may provide a measure that removes noise from the corresponding index of differential reactivity

more effectively than the untransformed measure of tonic arousal.

To explore this possibility, the observed sign of the difference between comparison and relevant questions was applied to the tonic measure prior to adding the tonic measure to the regression equation. Given a strong nonlinear relationship between tonic arousal and differential reactivity, large positive indices of differential reactivity would be associated with large positive measures of tonic arousal, and large negative indices of differential reactivity would be associated with large negative measures of tonic arousal. That is, the transformation would result in a strong linear relationship between tonic arousal and differential reactivity. As noted above, for suppression to occur, there should be a strong correlation between the suppressor variable and the predictor variable.

Table 5. *Proportions of Variance in the Criterion Explained by Indices of Differential Reactivity (R^2) and Increments in Variance Explained by Tonic Measures of Arousal (ΔR^2)*

		Entire Sample (N = 336)		*Probable-lie Effective Feedback (n = 60)*		*Directed Lie Effective Feedback (n = 60)*	
Base Model	Tonic Measure	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
SC Amplitude CP Amplitude Respiration Length	SC Level	.406**	.002	.633**	.019	.404**	.010
SC Amplitude CP Amplitude Respiration Length	SC Number of Responses	.406**	.001	.633**	.020	.404**	.006
SR Amplitude CP Amplitude Respiration Length	SR Level	.406**	.003	.622**	.004	.421**	.010
SC Amplitude SBP Amplitude Respiration Length	SBP Level	.423**	.002	.618**	.009	.423**	.004
SC Amplitude DBP Amplitude Respiration Length	DBP Level	.420**	.000	.597**	-.005	.415**	.028
SC Amplitude CP Amplitude Respiration Length HP Increase	HP Level	.408**	.019**	.633**	.014	.408**	.024

** $p < .01$

Table 6 shows the results of the multiple regression analyses when the transformed values of tonic arousal were included in the regression equation to predict the criterion. Since none of the increments in the proportion of variance explained was significant, there was no evidence that any transformed measure of tonic arousal increased the diagnostic validity of the base model.

Table 6. *Proportions of Variance in the Criterion Explained by Indices of Differential Reactivity (R^2) and Increments in Variance Explained by Transformed Tonic Measures of Arousal (ΔR^2)*

		Entire Sample (N = 336)		*Probable-lie Effective Feedback (n = 60)*		*Directed Lie Effective Feedback (n = 60)*	
Base Model	Tonic Measure	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
SC Amplitude CP Amplitude Respiration Length	SC Level	.406**	.001	.633**	.003	.404**	.000
SC Amplitude CP Amplitude Respiration Length	SC Number of Responses	.406**	.003	.633**	.000	.404**	.001
SR Amplitude CP Amplitude Respiration Length	SR Level	.406**	.007	.622**	.000	.421**	.009
SC Amplitude SBP Amplitude Respiration Length	SBP Level	.423**	.005	.618**	.002	.423**	.000
SC Amplitude DBP Amplitude Respiration Length	DBP Level	.420**	.002	.617**	.004	.415**	.033
SC Amplitude CP Amplitude Respiration Length HP Increase	HP Level	.408**	.007	.633**	.008	.408**	.017

** $p < .01$

Analysis of Covariance

A final attempt was made to test if tonic measures of skin conductance can be used in combination with indices of differential reactivity to improve the accuracy of polygraph outcomes. When the raw measurements of skin conductance response amplitude for a given individual are transformed to *z* scores, the resulting *z* scores have a mean of zero and unit variance. Since the *z* scores for all subjects have the same mean and the same variance, there are no differences among subjects in the average magnitude of skin conductance responses to test questions. Another method for removing differences among subjects is to use the original raw measurements of skin conductance amplitude and include the tonic level as a covariate in the regression equation.

Two measures of tonic skin conductance activity were derived. Therefore, two regression equations were created for each sample. The equations included raw differences between comparison and relevant response amplitude (in μS) and either log skin conductance level or log number of spontaneous skin conductance responses as the covariate. Table 7 summarizes the results of the regression analyses. To permit comparisons with the traditional approach, the regression equations that included indices of differential reactivity for skin conductance, cardiograph, and respiration based on differences between *z* scores are shown in the top row of Table 7. These results also appear along the top row of Table 6.

In the entire sample, all variables contributed significantly to their respective regression equations. As expected, the regression coefficients for both SC covariates (SC level and SC number of responses) were significant. These findings suggest that the covariates functioned as predicted; they removed variance among individuals in the raw differences in SC responses to comparison and relevant questions. However, the overall proportions of variance in the criterion explained by the current model with *z*-transformed measures of SC amplitude (R^2 = .406, .633, and .404) were generally greater than those obtained for either of the proposed models with raw measures of SC amplitude. Therefore, there was no advantage in using tonic SC level or the number of spontaneous responses as a covariate in the regression equation.

Although the proportions of variance explained by the regression models for the probable-lie test were generally greater than those for the directed lie test, none of the differences were statistically significant. Differences between probable-lie and directed lie tests are discussed in the final report for the original study (DoDPI97-P-0016).

Table 7. *Regression Results for Current Model with Standardized Measurements and Alternative Models with Absolute Measures of SCR Amplitude and Tonic Activity*

	Entire Sample (N = 336)		*Probable-lie Effective Feedback (n = 60)*		*Directed Lie Effective Feedback (n = 60)*	
Model	*B*	R^2	*B*	R^2	*B*	R^2
Standardized Measures with No Covariates		.406**		.633**		.404**
Standardized SCR Amplitude	.550**		.608**		.591**	
Standardized CPR Amplitude	.099*		.141		.086	
Standardized Respiration Excursion	.176**		.293**		-.025	
SC Level as a Covariate		.341**		.515**		.380**
SCR Amplitude (in □S)	.186**		.174		.128	
Standardized CPR Amplitude	.195**		.241*		.140	
Standardized Respiration Excursion	.187**		.301**		-.007	
SC Level (in μS)	.295**		.325*		.440*	
Number of Spontaneous SC Responses as a Covariate		.325**		.474**		.335**
SCR Amplitude (in □S)	.231**		.274*		.293	
Standardized CPR Amplitude	.221**		.308**		.209	
Standardized Respiration Excursion	.183**		.286**		-.028	
SC Number of Responses	.226**		.154		.220	

*Differed from 0, $p < .05$.
**Differed from 0, $p < .01$.

Discussion

One objective of the present research was to assess relationships between tonic levels of arousal and the magnitude of phasic reactions to test questions. Measures of tonic arousal were obtained from 336 subjects who participated in a previous experiment (DODPI97-P-0016). That experiment was designed to assess the effects of the demonstration test on the accuracy of subsequent probable lie and directed lie polygraph examination. It also assessed the effects of various types of feedback concerning the outcome of the demonstration test on the accuracy of the subsequent polygraph examinations. Preliminary tests for heterogeneity of regression revealed no evidence that the strength or direction of relationships between tonic arousal and phasic reactivity varied over treatment conditions. Analyses of group means revealed several significant effects of Feedback and interactions of Feedback with other factors on selected physiological measures. However, the effects were generally small, and they did not affect the nature of the relationships between tonic and phasic measures of autonomic activity.

These preliminary tests justified an analysis of all subjects. Nevertheless, in addition to analyzing the entire sample, we conducted parallel analyses of the probable-lie and directed lie groups that had been given numbers tests and effective feedback. The two effective feedback groups were analyzed separately because the procedures for those groups were most similar to the procedures used by field polygraph examiners.

Six measures of tonic arousal were derived. They included the number of nonspecific skin conductance responses, skin conductance level, skin resistance level, systolic and diastolic blood pressure levels, and heart period level. The results obtained for the entire sample were similar to those obtained for the effective feedback subgroups. For the entire sample, five of the six measures of tonic arousal correlated significantly with the mean magnitude of phasic responses to test questions. However, the correlations for the cardiovascular measures were generally small and were statistically significant only for the entire sample. This suggests that there may be small but reliable relations between tonic and phasic measures of cardiovascular activity, but the power was too low to provide evidence of these relationships in the effective feedback subgroups.

Phasic skin conductance responses correlated highly with the number of nonspecific skin conductance responses ($r > .66$) and the skin conductance level ($r > .57$). These findings are consistent with those reported by Hord et al. (1964) and by Schell, Dawson, and Filion (1988). Hord et al. reported correlations between skin conductance level and skin conductance responses that ranged from .35 to .77.

Skin resistance was derived from measured skin conductance activity. The expected negative correlation between tonic skin resistance and skin resistance was significant for the entire sample. However, the correlation was considerably smaller in magnitude ($r = -.20$) than the correlations obtained for skin conductance, and the correlation was not significant for either of the effective feedback groups. These results suggest that skin resistance responses are less dependent on basal levels of activity than are skin conductance responses.

Hord et al. (1964) also reported strong negative correlations between heart rate level and increases in heart rate. In their study, subjects with high tonic heart rates showed the smallest increases in heart rate. In the present study, heart period level was positively correlated with increases in heart period but only marginally so. Since heart period and heart rate are inversely related, a positive correlation between measures of tonic and phasic heart period was expected.

However, the magnitude of the correlation obtained in the present study ($r = .14$) was considerably less than the correlations reported by Hord et al. ($r = -.36$ to $-.64$). This discrepancy may be due in part to the nonlinear nature of the relationship between heart period and heart rate. To derive heart rate (in BPM) from heart period (in seconds), one finds the reciprocal of the heart period and multiplies by 60. Although there is a strong negative relationship between heart

rate and heart period, it is not perfect. When we reexamined the tonic-phasic relationship after transforming our measures of heart period to heart rate, the strength and sign of the correlation changed from r = .14 to r = -.21. The sign of the correlation for heart rate was consistent with Hord et al., and the magnitude of the correlation increased. However, the observed relationship was still substantially weaker than those reported by Hord et al..

The present study also differed from the study by Hord et al. in how the cardiac response was defined. Hord et al. asked subjects to listen for the onset of a tone and measured *increases in heart rate.* In contrast, we measured *increases in heart period*, which correspond to decreases in heart rate. Measures of cardiac deceleration are traditionally used in research on polygraph techniques (Podlesny & Kircher, 1999; Raskin, 1979) and are more diagnostic than measures of cardiac acceleration (Kircher & Raskin, 1988). The drop in heart rate in response to test questions with signal value may be indicative of an orienting response (Graham & Clifton, 1966), or it may be a reflexive response to a rapid rise in blood pressure (Raskin, 1979).

The availability of continuous measures of blood pressure and the large sample size provided an opportunity to detect even small correlations between tonic levels of blood pressure and phasic blood pressure reactions to test questions. The correlation between tonic diastolic blood pressure and the phasic diastolic blood pressure response was statistically significant for the entire sample. However, the correlation was small (r = .11), and it was not significant for either effective feedback group. In addition, the sign of the correlation was negative for one subgroup and was positive for the other. There was no evidence of a relationship between tonic and phasic systolic blood pressure responses. These results suggest that if there is a linear relationship between tonic and phasic blood pressure activity during polygraph examinations, then it is a small one and, for all practical purposes, may be ignored.

Current computer and numerical scoring methods for analyzing polygraph charts do not consider individual differences in tonic levels of physiological activity. Decisions are based exclusively on differences between phasic reactions to comparison and relevant questions. Not only are individual differences in tonic arousal ignored, at least two computer programs systematically remove differences among individuals in phasic responses as well (Kircher & Raskin, 1988; Olsen et al., 1997). These algorithms transform raw measurements of reactions to comparison and relevant questions to standard scores. The transformation to standard scores is a linear one so that a subject who reacts more strongly to a particular type of question (e.g., relevant) will continue to show proportionally stronger reactions to that type of question. However, since the standard scores for all subjects have the same mean (M_Z = 0) and variance (S^2 = 1), the transformation removes all differences among subjects in the mean magnitude and variance of responses to the test questions. Although the advantages of within-subject standardization and within-subject comparisons of reactions to comparison and relevant questions have been documented (e.g., Kircher & Raskin, 1981; 1988; Podlesny & Kircher, 1999; Raskin, Kircher, Honts, & Horowitz, 1988), it was not known if additional diagnostic information could be extracted from absolute levels of tonic arousal. The primary purpose of the present study was to explore that possibility.

Three of the six measures of tonic autonomic arousal distinguished between guilty and innocent subjects. Consistent with expectations, in the entire sample, skin resistance was lower and heart rates were higher for guilty subjects than for innocent subjects. Together, these results suggest that guilty subjects were more highly aroused during their polygraph examinations than were innocent subjects. However, the effect on skin resistance was small (r_{pb} = .11) and was not significant for the probable-lie or directed lie subsamples. In addition, the effect on heart period was substantial only for probable-lie subjects (r_{pb} = .42). For directed lie subjects, the effect was not significant (r_{pb} = .09). A third, marginally significant difference between guilty and innocent subjects in the probable-lie condition was obtained for systolic blood pressure level (r = -.27). The guilty subjects in that condition had higher tonic

systolic blood pressure than did innocent subjects.

The observed, weak relationships between measures of tonic arousal and the criterion were not surprising. There are large differences among individuals in levels of basal autonomic activity (Sternbach, 1966). Having committed a mock crime, subjects' tonic levels of autonomic activity increased. However, with the exception of heart rate for probable-lie subjects, the increases were small relative to the wide range of individual differences inherent in these measures.

Despite the small effects of deception on tonic measures of arousal, it was still possible that they would improve the accuracy of polygraph decisions when combined with traditional measures of differential reactivity. Multiple regression was used to test for an increase in the proportion of criterion variance explained when measures of tonic activity were used in combination with measures of differential reactivity. The proportion of criterion variance explained and the expected proportion of correct decisions are monotonically related. However, the proportion of criterion variance explained was chosen as the primary outcome measure because it is a more sensitive index of predictive validity than the accuracy of decisions.

Initially, a regression equation was created using only measures of differential reactivity to comparison and relevant questions. A measure of tonic arousal was then added to the equation, and its contribution to the proportion of criterion variance explained was tested for significance. To facilitate interpretation of the results, only one tonic measure was added to the regression equation at a time.

A tonic measure of arousal could contribute to the regression equation in one of two ways: as a *predictor* variable or as a *suppressor* variable (Cohen & Cohen, 1975). In general, a useful *predictor* variable accounts for variance in the criterion that is not already explained by other variables in the regression equation. The most useful predictor variables are those that are highly correlated with the criterion and are relatively independent of other variables in the equation.

A *suppressor* variable, on the other hand, contributes to the regression equation by removing (suppressing) a portion of variance from a predictor variable that is unrelated to the criterion. A suppressor variable filters noise from predictor variable and contributes to the regression equation indirectly, by increasing the correlation between the predictor variable and the criterion. The most useful suppressor variables are those that are uncorrelated with the criterion but are highly correlated with one or more predictor variables.

Except for heart period, none of the tonic measures correlated with the criterion. Therefore, there was little hope that any of the tonic measures would serve to predict the criterion directly. There was, however, still some chance that one or more tonic measures of arousal would contribute to the regression equation indirectly by serving as suppressor variables.

Three attempts were made to determine if any of the six tonic measures of arousal would contribute significantly to a base model. The base model contained indices of differential reactivity to comparison and relevant questions for electrodermal, cardiovascular, and respiration channels. The composition of the base model was varied to ensure that each tonic measure was added to a regression model that contained an index of differential reactivity derived from the same channel. For example, when testing the contribution of tonic skin conductance level, the electrodermal index of differential reactivity in the base model was derived from skin conductance responses. Conversely, when testing the contribution of tonic skin resistance level, the electrodermal index of differential reactivity was based on skin resistance responses. This approach was based on the assumption that a tonic measure of arousal would correlate more highly with an index of differential reactivity derived from the same physiological signal than from any other signal. By maximizing the correlation between tonic arousal and an index of differential reactivity, we hoped to increase the chances that the tonic measure would suppress

variance in the index of differential reactivity and improve its ability to predict the criterion.

In the first set of analyses, the original measures of tonic activity were entered into the regression equation. In the second set of analyses, transformed measures of tonic activity were entered into the regression equation. Tonic measures were transformed by assigning the sign of the associated index of differential reactivity to the tonic measure before adding the tonic measure to the regression equation. The final set of analyses tested only the contributions of tonic skin conductance level and the number of spontaneous skin conductance responses. In those analyses, differences between the reactions to comparison and relevant questions measured in the original units (μS) were used in the regression equations rather than differences between standardized measurements of reactions to comparison and relevant questions.

The results of these tests were generally disappointing. In only one case did the measure of tonic arousal contribute significantly to the proportion of variance explained. In the analysis of the entire sample of 336 subjects, heart period contributed significantly to the prediction equation. However, heart period accounted for less than 2% of the variance in the criterion and did not improve the accuracy of decisions.

Conclusions

It is difficult to draw firm conclusions from null results when the sample size is small. However, when the sample size is as large as it was in the present study, estimates of population parameters are quite stable. Under these circumstances, it may be concluded that if there are any benefits in using tonic measures of arousal for diagnosing truth and deception, they are likely to be minimal. Based on the present results, there appears to be little need to call on polygraph manufacturers to develop instrumentation that is capable of providing absolute measures of tonic physiological activity.

It is important to note that the data for the present study were obtained from subjects who participated in a mock crime experiment. Whether these laboratory results are representative of the field is an open question. Truthful and deceptive field suspects may show higher levels of tonic arousal than subjects in a laboratory experiment. Indeed, differences in tonic arousal may mediate the observed variance among laboratory studies in the accuracy of polygraph outcomes (Kircher, Horowitz, & Raskin, 1988). However, if the effects of setting were constant for truthful and deceptive subjects, one would not expect measures of tonic arousal to be any more useful in the field than in the lab. We previously compared data from our lab and field polygraph studies and found that differences between comparison and relevant questions in the field sample were generally shifted in the negative direction (Kircher, Raskin, Honts, & Horowitz, 1994). The truthful and deceptive field suspects appeared more deceptive on their polygraph tests than did the truthful and deceptive laboratory subjects. Although the differences between comparison and relevant questions were more negative, the separation between truthful and deceptive individuals was similar for the lab and field samples. The various indices of differential reactivity were as diagnostic for the field sample as they were for the lab sample. In addition, the variances and covariances among various indices of differential reactivity in the lab and field samples were indistinguishable. Thus, the field suspects may have been more physiologically aroused than the laboratory subjects, but this increase arousal had no discernable effect on the diagnosticity of the physiological measures or their interrelationships. The findings from the Kircher et al. study suggest that our laboratory procedures produce relationships between indices of differential reactivity and the criterion and among indices of differential reactivity that closely approximate the relationships that occur in the field. To the extent that they do, the present results suggest that tonic measures of arousal would not prove useful for field applications.

References

Bell, B.G., Raskin, D.C., Honts, C.R., & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph, 28,* 1-9.

Boucsein, W. (1992). *Electrodermal Activity.* New York: Plenum Press.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* New York: Wiley.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Graham, F. K., & Clifton, R. K. (1966). Heart rate change as a component of the orienting response. *Psychological Bulletin, 65,* 305-320.

Hord, D. J., Johnson, L.C., & Lubin, A. (1964). Differential effect of the law of initial values (LIV) on autonomic variables. *Psychophysiology, 1,* 79-87.

Katkin, E.S. (1975). Electrodermal lability: A psychophysiological analysis of individual differences in response to stress. In I. G. Sarason & C. D. Spielberger (Eds.) *Stress and anxiety* (Vol.2). Washington, D. C.: Hemisphere.

Kircher, J. C., Raskin, D. C., Honts, C. R., & Horowitz, S. W. (1994). Genereralizability of statistical classifiers for the detection of deception. *Psychophysiology, 31,* S73. (Abstract)

Kircher, J. C., & Raskin, D. C. (1981). Computerized decision-making in the detection of deception. *Psychophysiology, 18,* 204-205. (Abstract)

Kircher, J.C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291-302.

Kircher, J.C., Woltz, D.J., Bell, B.G., & Bernhardt, P.C. (1998). *Effects of audiovisual presentations of test questions during relevant-irrelevant polygraph examinations and new measures.* Final report to the Central Intelligence Agency (Grant No. 110200-997-MO). Salt Lake City: University of Utah, Department of Educational Psychology.

Krapohl, D., & Ansley, N. (1999, August) *Prevalence of diagnostic features in polygraph tracings.* Paper presented at the annual meeting of the American Polygraph Association, Dallas, TX.

Olsen, D. E., Harris, J. C., Capps, M. H., & Ansley, B. A. (1997). Computerized polygraph scoring system. *Journal of Forensic Sciences, 41*(1), 61-70.

Podlesny, J. A. & Kircher, J. C. (1999). *The Finapres (volume clamp) recording method in psychophysiological detection of deception examinations: Experimental comparison with the cardiograph method.* Final report to the National Institute of Justice (91-IJ-R-026). Laboratory Division, Federal Bureau of Investigation.

Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In K.D. Kimmel, E. H. Van Olst & J. F. Orlebeke (Eds.) *The orienting reflex in humans.* Hillsdale: Lawrence Erlbaum Associates.

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.) *Psychological methods in criminal investigations and evidence* (pp 247-296). New York: Springer Publishing.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35,* 651-665.

Rushmer, R. F. (1976). *Organ physiology: Structure and function of the cardiovascular system (2nd Ed).* Philadelphia: Saunders.

Schell, A. M., Dawson, M. E., & Filion, D. L. (1988). Psychophysiological correlates of electrodermal lability. *Psychophysiology, 25,* 619-632.

Spielberger, C. D., Gorsuch, R.L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA.: Consulting Psychologists Press.

Sternbach, R. A. (1966). *Principles of psychophysiology.* New York: Academic Press.

Venables, P. H., & Christie, M. J. (1980). Electrodermal activity. In I. Martin & P. H. Venables (Eds.) *Techniques in psychophysiology.* New York: John Wiley & Sons.

Wilder, J. (1967). *Stimulus and response: The law of initial value.* Bristol: Wright.

The Infamous James Alphonso Frye[1]

Brett A. Stern & Donald J. Krapohl[2]

Abstract

Among lawyers and polygraph practitioners, the case of *Frye v. United States,* is one of the most well-known legal cases involving the admissibility of novel scientific evidence and, more specifically, the admissibility of polygraph evidence. This paper does not address the legal merits of the Frye case, rather it explores many of the myths surrounding the polygraph field's most infamous examinee. This exploration is undertaken from review of the literature and the perspective of several influential people who helped shape the case for and against

On November 27, 1920, a 55-year old prominent physician lay face down and dead at 1737 Eleventh Street, N.W., Washington, D.C., the victim of a fatal gun shot to the head. Two of five bullets stuck the physician; one entered his " . . . right jaw and passing through the left side of his head and another bullet lodging in the right armpit. Three bullets went wide of their mark" (Washington Post, 1920, November 28). The Washington Post (1920, November 28) also reported that according to an eyewitness, four shots were fired. Other reporting states that Dr. Brown was " . . . shot twice in the breast and once in the head" (Excerpt from an article entitled, "*Negro Held on Charge of Slaying Physician . . .*" ; article appeared in an unidentified and undated newspaper). Even the man ultimately convicted of the crime, James A. Frye, apparently reported in a confession he later recanted, that he was not sure whether he fired four or five shots (Evening Star, 1921 August 23). Notwithstanding the various media reports, there is no dispute that Dr. Brown suffered a mortal gun shot wound to the head. So begins the legacy of the now infamous James Alphonso Frye. Perhaps no federal legal case involving the issue of "lie detection" has captured the attention and imagination of so many as *Frye v. United States* (1923). At Frye's trial, " . . . unprecedented excitement surrounded the debut of the mystical machine that could ferret out truth from deception for the courtroom was crowded to overflowing throughout the morning by persons anxious to see the sphygmomanometer tested" (Excerpt from an article entitled, "*Holds Frye Guilty of Killing Doctor . . .* "; article appeared in an unidentified newspaper in 1922).

The Frye case took on legendary proportions and without question significantly impacted jurisprudence in this country—particularly with respect to the standard of admissibility of novel scientific evidence. However, at issue remains the question, was James A. Frye guilty of killing Dr. Robert Wade Brown in his home on that inauspicious Saturday in November? As newspaper articles covering the murder pointed out, Dr. Brown was not only a " . . . prominent colored physician" he was " . . . the best known colored physician in the city" and was also " . . . president of the National Benefit Insurance Company {NBIC}" (Washington Post, 1920, November 28). Dr. Brown's family offered a $1,000.00 reward, the police offered a $100.00 reward, and Dr. Ferdinand Whitby, a friend of Dr. Brown, offered another $50.00 for information leading to the arrest of the killer (Evening Star, 1921, August 23).

---

[1]This article is one in a series under the heading Polygraph Myths. The opinions expressed in this article are exclusively those of the authors, and do not necessarily represent those of the Department of Defense or the U.S. Government.

[2]Brett A. Stern, Senior Instructor, Instruction Division, Department of Defense Polygraph Institute; Donald J. Krapohl, Research Division, Department of Defense Polygraph Institute.

It was also reported (Excerpt from an article entitled: "*Dr. Brown Killed*"; article appeared in an unidentified newspaper bearing no date) that NBIC offered a $1,000.00 reward, but it is unclear whether the Brown family reward and the NBIC reward were one and the same.

The Myths Abound

Among the more contentious myths surrounding James Frye is that he was innocent of killing Dr. Brown, (sometimes reported as a white man [Jasanoff, 1989] or correctly as a black man) and that Dr. William Marston proved Frye's innocence through his discontinuous systolic blood pressure test (Marston, 1938). An equally contagious myth is the one advanced by polygraph insiders (Marston, 1938; Matte, 1980) and others in the legal profession and academia (August, 1975; Jasanoff; 1989, Marino, 1975) claiming Frye was convicted of murder and subsequently released when the actual killer came forward. For example, Matte wrote, "Three years later (after his conviction) Frye was freed as a result of further investigation, . . . " August wrote: "The defendant, after serving three years of a life sentence, was exonerated by the confession of the actual killer, and released." Jasanoff claimed although Frye was found guilty of second-degree murder he " . . . was subsequently exonerated when another man confessed to the crime."

Another myth mostly promulgated by Dr. Marston (1938), is that Frye's confession, which he later recanted, stemmed from a promise made to him by an unidentified friend who said that he would fix it so Frye got half the reward for his own conviction and a full pardon soon after. Moreover, Dr. Marston (1938) alleged that the person who duped Frye into confessing was the real murderer.

The Jury Verdict

The Washington Post (1922, July 29) article headline read: "*Life-Sentence Penalty in Murder of Doctor.*" Though he was indicted for murder in the first degree, Frye escaped a death sentence when the jury returned its verdict of guilty of murder in the second degree in violation of Section 23, Title 6, D.C. Code (*U.S. v. Frye*; Docket Entries, n.d.). The jury received the case at 3:30 p.m. and by 4:30 p.m. that same day, it completed its deliberations and rendered its verdict (Excerpted article, 1922).

If one were inclined to believe Dr. Marston (1938), it was his exculpatory "lie detector" test that moved the jury to convict on the charge despite the fact that exculpatory test results were never admitted into evidence. Dr. Marston apparently dismissed the possibility that the trial judge's instructions may have been equally as persuasive as his "lie detector" test in saving Frye's life.

Judge Walter I. McCoy, presiding judge, provided the defendant and jury with 12 instructions. Instruction number 12 essentially set forth that because premeditation had not been shown, the jury could not bring in a guilty verdict of murder in the first degree. Irrespective of what motivated the jury to return the verdict it did, " . . . Frye was the happiest colored boy in Washington that night," according to Dr. Marston (1938). Notwithstanding the judge's instructions, Frye and others give us reason to ponder the wisdom of the jury and the U.S. Government's position.

The U.S. Government Takes a Stand

About thirteen years after his arrest, the U.S. Government was no less certain of Frye's culpability than the jury that convicted him. In response to Frye's Application for Executive Clemency, Leslie C. Garnett (1934), United States Attorney, wrote, "There is no doubt in my mind that he is guilty of the crime of which he was convicted. I think he was given all the leniency to which he was entitled when the jury returned a verdict of second degree murder."

The Legacy Continues

Irrespective of whether the myths surrounding James Frye's culpability in the death of Dr. Brown ever fade away, his name will live on in the annals of U.S. legal history. The Frye case continues to capture the attention of legal scholars. This is because Judge McCoy's decision barring the introduction of Dr. Marston's "lie detector" test, and the Appeals Court's decision set forth a precedent in U.S. jurisprudence that some courts continue to grapple with today. While most courts follow the more recent rulings in the trilogy of U.S. Supreme Court decisions in

Daubert v. Merrill Dow Pharmaceuticals, Inc (1993); *General Electric Co. et al. v. Joiner et ux.* (1977); and *Kumho Tire Co., LTD., et al., v. Carmichael et al.* (2000), some courts continue to rely exclusively on the "general acceptance" standard outlined in *Frye v. United States* (1923).

The Search for Frye

"Doctor Shot Dead; Assailant Flees . . . " is what The Washington Post headline article read on Sunday, November 28, 1920, the day after Dr. Brown was slain. The D.C. police were initially searching for James Frye not for killing Dr. Brown but for the highway robbery of G. R. Blake, an automobile salesman of Indianapolis, Indiana who was visiting the D.C. area. Moreover, the U.S. Secret Service (USSS) was looking for Frye for Treasury check fraud (Bratton, 1921; Washington Post, 1921, August 29).

USSS Weighs In

The USSS first became aware of James Frye in connection with Treasury check fraud, when the victim, Irving Q. Fields and Dr. John R. Francis (local dentist), named Frye as a possible suspect (Bratton; McCahill, 1921, August 16). Shortly thereafter, Inspector Grant, D.C. Police Department, telephoned Operative McCahill, USSS and informed him that they (Detectives Jones and Jackson of the central office) had arrested Frye for robbery. McCahill went to D.C. Police Headquarters and arrested Frye on August 18, 1921 for the forgery of a $77.50 check (check number 830308, dated June 30, 1921) issued to Irving Q. Fields.

Suspicion focused on Frye when he failed, on two accounts, to properly endorse a check he tried to pass. The first time he endorsed the check he signed it Irvin instead of Irving and the next attempt he used the middle initial "O" instead of "Q." McCahill believed that Frye was the probable forger when he compared Frye's handwriting, which he obtained from the D.C. police, with the writings on the check. McCahill (1921, August 24) interviewed Frye at D.C. Police Headquarters, obtained another handwriting sample and made a positive match to the forgery. Frye subsequently confessed to the forgery. It is unclear whether Frye served any period of incarceration specifically for the forgery.

Shortly following Frye's forgery confession, Inspector Grant, D.C. Police Department, told McCahill that an unidentified man came in and told him that Frye murdered Dr. Brown. McCahill (1921, August 24) continued to question Frye and told him that he would forget the forgery charge if Frye agreed to tell Inspector Grant about the robbery of G.R. Blake and murder of Dr. Brown.

McCahill's inducement for Frye to cooperate was addressed in jury instruction number nine, at Frye's subsequent trial. Instruction nine essentially said if " . . . the defendant was influenced by any hope of reward or fear of punishment . . . however slight . . . it is your duty to reject the statement." Frye initially denied the murder but admitted to the robbery (Evening Star, 1921, August 23). After a few more days of questioning by Inspector Grant and others, Frye admitted to murdering Dr. Brown. Frye's confession was initially recorded stenographically and then it was typewritten and signed by Frye in the presence of witnesses (*U.S. v. Frye*, Bill of Exceptions). The trial record reflects that the jury was not persuaded that Frye was offered any inappropriate inducement to confess.

Dr. Marston's Account

It would be remiss to write about the Frye case and not address Dr. William Moulton Marston's (originator of the "lie detector" [Larson, 1938]) opinion regarding Frye's culpability. Dr. Marston was a central figure in the case, believing he literally saved Jim's (as Dr. Marston called him) life. Marston (1935) reported that Frye and his attorneys " . . . gave the lie detector full credit for saving him from otherwise certain hanging." One can't help get the impression that Marston (1938) took a liking to Frye for he characterized Jim as " . . . a good-natured likeable young Negro, who had gotten into trouble by associating with the wrong crowd, a gang of toughs."

According to Marston (1938), D.C. police detectives, while investigating the robbery of G. R. Blake, focused on two men,

Bowie (Bratton) and Frye. Frye and his accomplices (Bowie and another unidentified man) robbed Mr. Blake of a diamond ring reportedly valued at $1,500.00, a watch, and $75.00 in cash. When Frye was arrested for robbery he had Blake's watch in his possession (Excerpted article "*Negro Held ...*") . He received a sentence of four years for the robbery (Garnett, 1934).

During the robbery investigation Frye was also questioned about Dr. Brown's murder but " . . . denied all knowledge of it" (Marston, 1938). Unable to post bail, Frye was held in jail. Frye subsequently confessed to Inspector Grant and detectives that he murdered Dr. Brown (Evening Star, 1921, August 23). The Washington Post (1921, August 23) reported that Frye, while in jail, provided a written confession to Inspector Clifford Grant, Detectives Paul W. Jones and John T. Jackson correctly detailing the caliber of the weapon used in the killing, how the killing took place, how he made his escape, and where he hid the pistol. Was Frye's detailed revelation of the murder of Dr. Brown just a coincidence? Was Frye the real killer or was he part of a conspiracy gone awry?

The question that begs to be answered is if Frye was not responsible for killing Dr. Brown, why would he confess to doing so? The answer may lie in words Frye allegedly told his two court appointed attorneys when they met. According to Marston (1938), Jim told his attorneys, "I am innocent! I had nothing to do with Brown's murder. I confessed because I was promised half the reward for my own conviction."

According to Marston, Frye's attorneys, Richard Mattingly and Lester Wood (both court appointed, as Frye was declared legally indigent) wanted Marston to administer a "deception test" to Frye, not to exculpate him, but rather to show Frye the futility of his assertion of innocence (Marston, 1938). Doctor Marston administered his test to Frye in the District jail, the results of which, according to Marston, showed that "Frye's final story of innocence was entirely truthful! His confession to the Brown murder was a lie from start to finish" (Marston, 1938). Marston further asserted Frye's false confession was motivated because " . . . a certain Negro friend . . . had assured him that he would fix it so he got half the reward for his own conviction and a full pardon soon after." Marston (1935) reported that his "lie detector" also verified Frye's underlying reason for his false confession. Who was this Negro friend? Marston tells us that " . . . Frye had been approached by a Negro who was supposed to be head of a drug ring and . . . that he (the Negro) would have profited by Dr. Brown's death." Marston also tells us that the Negro friend, " . . . had, in fact officially turned Frye in as the murderer, and had claimed the reward."

With not a single witness available to corroborate Frye's account of his false confession, and fearing a murder one conviction, Frye's attorneys offered to qualify Dr. Marston as an expert for the purpose of introducing Frye's exculpatory test results (Marston, 1938). Marston (1935) asserted the offer was made in front of the jury and even though the evidence was not admitted the jurors were moved by the proffer. "The test undoubtedly saved his life. No jury could help being influenced by the knowledge that Frye's story had been proved truthful by the 'lie detector'," according to Marston (1938). Instead of being hanged, Frye was sentenced to imprisonment for life, though he served only two months short of 18 years (Starrs, 1983). Marston (1935) stated that " . . . the jury acquitted him (Frye) of first degree murder" and that "He was sentenced on another charge, giving time for investigation which verified the lie detector findings."

Notwithstanding Dr. Marston's comments, Frye appealed his conviction on the basis that the court erred on eight accounts, five of which specifically dealt with the court's failure to recognize Dr. Marston's expertise in detecting deception and the court's unwillingness to allow the introduction of Dr. Marston's evidence (Starrs, 1983; *U.S. v. Frye*, Case File 38325). The Court of Appeals of District of Columbia addressed only one Assignment of Error; specifically, that the " . . . trial court erred by refusing to allow an expert witness to testify as to the result of a systolic blood pressure deception test taken by appellant" (*U.S. v. Frye*, case file 38325).

Marston asserted in his published works that " . . . further investigation showed that the Negro who had tried to put Jim on the spot by inducing him to make a false confession was Dr. Brown's real murderer" (Marston, 1938). Marston failed to fully name the person he believed was responsible for Dr. Brown's murder other than to intimate that " . . . we may call him J.W." Doctor Marston's claim that J.W. was the murderer, and author Matte's (1980) similar claim, perpetuated the myth that Frye was innocent.

The World According to Frye

Equally apropos to Dr. Marston's take on the case is Frye's position regarding the murder of Dr. Brown, and Frye's position is at times arguably plausible if not a bit too fanciful and convenient. In his confession, that he later recanted, Frye told Detective Grant and inspectors Jones and Jackson that he went to Dr. Brown's office to get a prescription filled. He had one dollar but Dr. Brown told him the cost to fill the prescription was two dollars. Frye offered his pistol as collateral for the other dollar. Dr. Brown did not accept the pistol as collateral. An argument apparently ensued and Frye claims that Dr. Brown pushed him to the floor. Frye fired four or five shots while down on the floor (Evening Star, 1921, August 23). Following his recantation of his confession, Frye was never again heard to have admitted killing Dr. Brown.

According to Dr. Marston, Frye recanted his confession because he realized he was being played as a fool by his supposed unidentified co-conspirator and alleged friend (Marston, 1938). Marston (1935) asserted that Frye retracted his confession because his co-conspirator was unable to collect the reward that he promised to split with Frye upon release. In his 1936 Application for Clemency, Frye made the following interesting statement, "I have seen my mistake in life and am only asking for a chance to prove to society I am worthy of being accepted once more as an American citizen." It is not known specifically to what mistake Frye was referring—Treasury check fraud, robbery, murder, or something else.

Frye named Dr. Brown's killer and offered corroboration—or did he? In his 16-page, somewhat convoluted attachment to his Application for Executive Clemency, Frye named Dr. John R. Francis as Dr. Brown's murderer. Frye asserted he and Francis were " . . . looked upon as friends . . . " and that they " . . . could have easily passed as brothers and very often were taken as such" (Application, 1936).

In a memorandum from USSS Operative Bratton (1921) to his USSS chief, he acknowledged that Frye and his co-conspirator in the robbery of Blake were indeed friends of Dr. Francis. Moreover, Bratton wrote, "From what I could learn by a discreet investigation, Doctor Francis is connected with many of the colored crooks in the neighborhood which is a bad one. This information must be kept strictly confidential." Was this merely a case of mistaken identity as Frye suggested? Frye claimed, while Francis was under the influence of some unidentified narcotic, which he allegedly purchased from a narcotics dealer while in Frye's presence, Dr. Francis told Frye that he (Francis) killed Dr. Brown, although the motive was never mentioned (Application, 1936).

Frye asserted that Dr. Francis was concerned that someone named Broaduax might finger him as the killer because Broaduax saw Dr. Francis run from Dr. Brown's office to his (Dr. Francis') office, presumably after the shooting (Application, 1936). However, we never learn more of this man named Broaduax and he was never identified or called as a witness at Frye's trial.

Frye contended that another unidentified man visited his aunt and told her that he knew Frye was not guilty of killing Dr. Brown but " . . . did not want to say anything for fear he would get himself into trouble" (Application, 1936). Frye did not elaborate further as to the identity of this potential savior or what he had to say.

According to Frye, another person, Lois Dunlap, supposedly knew Dr. Francis visited Dr. Brown's house on the day of the killing and Dr. Francis feared Dunlap saw him running back to his office after the killing because Dunlap was in Dr. Francis' house at the time (Application, 1936).

Who was Broaduax? Why wasn't Lois Dunlap called as a defense witness? What motivated the unidentified man to come forward on Frye's behalf yet never to be heard from again? How would his coming forward get him into trouble? Unfortunately, the answers to these questions will remain a mystery.

Was Frye the unfortunate victim of a conspiracy gone amiss? Frye contended Dr. Julian Jackson (who was at Dr. Brown's residence when Dr. Brown was shot to death) knew Frye was not the killer but was afraid to finger Dr. Francis as Brown's murderer (Application, 1936). According to Frye, only after a subsequent conversation with D.C. Metropolitan Police detective Paul W. Jones did Dr. Jackson finger Frye as Dr. Brown's killer (Application, 1936). Why was Dr. Jackson fearful of identifying Dr. Francis as the true killer? Frye never tells us.

Perhaps in an effort to scam "the system" Frye was duped by the police into confessing—at least that is what he would have us believe. Frye said he confessed to Dr. Brown's murder after Detective Jones propositioned him. Detective Jones, according to Frye, told him that he (Jones) " . . . could and would squash the robbery charge if I would accept his proposition." Detective Jones supposedly told Frye to confess to Dr. Brown's murder for he had a solid alibi that would ultimately vindicate him, and yet he would reap some of the proceeds of the $1,000.00 reward. On its face value, Frye's supposition seems too far-fetched to warrant any merit. However, there is evidence that parallels what Frye asserts. If one may recall, Operative McCahill of the USSS actually told Frye that he would drop the forgery charge if Frye agreed to tell Inspector Grant about the robbery and murder.

Could Frye be so misguided as to risk his life on the words of man, a police officer, who he never met? What would make Frye believe they would be entitled to half the reward money if they falsely confessed to Dr. Brown's murder?

Frye's principal defense was an alibi (Motion for Continuance, 1922). Frye's ace-in-the-hole was apparently Mrs. Essie Watson and Miss Marion L. Cox. In their Motion for Continuance (1922), Frye's attorneys, Mattingly and Wood, stated, "That we expect and believe that she {Essie Watson} will testify that it is the best of her knowledge and belief that defendant was a visitor at her home at the time when the murder is alleged to have been committed . . . " Were Watson and Cox the alibis Detective Jones was allegedly referring to who would ultimately set Frye free? Mrs. Watson was apparently too ill to testify at Frye's trial though her "death bed" statement, attesting to Frye's presence at her house at the time the killing took place, was reportedly read aloud in court, though the trial record does not reflect such a reading having taken place (Application, 1936).

Frye maintained he was visiting Miss Cox at Mrs. Watson's home at the time Dr. Brown was murdered. Frye stated in his clemency application that he and Miss Cox were " . . . on very friendly terms" and they had a subsequent quarrel " . . . that I suppose befall every couple" (Application, 1936). Again, bad luck befell Frye for neither Mrs. Watson nor Miss Cox ever appeared as a witness at Frye's trial. Frye claimed the man who was later to become her husband prohibited Miss Cox from testifying on Frye's behalf (Application, 1936).

The Star Witness

Dr. Julian D. Jackson's account of Dr. Brown's killing, as reported by The Washington Post, was arguably the most persuasive in light of the jury's verdict. Dr. Jackson, of Norfolk, Virginia was visiting at Dr. Brown's house at the time Brown was killed. Dr. Jackson admitted the would-be killer into Brown's house just after 8:30 p.m. and shortly thereafter, while sitting in the kitchen, heard four gunshots (Washington Post, 1920, November 28). Dr. Jackson reportedly told detectives he heard the supposed patient say: "Did you do anything for me?" Dr. Brown answered in the negative. Soon thereafter, Dr. Brown lay dead of a fatal gun shot wound to the head. Jackson ran to the front door as the killer made his get-away by jumping over the front yard fence. As Jackson stood on Dr. Brown's front porch, " . . . the Negro shot at Jackson." Dr. Jackson later identified Frye as Dr. Brown's killer and the

man who shot at him during his escape (Washington Post, 1920, November 28).

Just the Facts

Leslie C. Garnett was the U.S. Attorney for the District of Columbia. She wrote in a letter to the Attorney General (July 21, 1934) a response to Frye's Application for Clemency wherein she left no doubt concerning the government's certainty about Frye's culpability and the leniency the jury showed him. In that same letter she outlined the facts of the case. That letter read in substantial part as follows:

> About 5:30 p.m., on November 25, 1920 {this date is factually in error} the defendant went to the office of Dr. Robert W. Brown, the deceased. Dr. Julian D. Jackson, who was visiting Dr. Brown, answered the door and told defendant that Dr. Brown was not in. So defendant went away and returned about 8:45 p.m. Another man named William Robinson arrived at the same time as defendant did, and defendant told Robinson to go in first, as he was in no hurry, and Robinson did so. When Robinson came out of the doctor's office he noticed that the defendant had put on a pair of smoked glasses. Then Robinson went away and defendant entered the office of Dr. Brown. Dr. Jackson had let defendant in the second time and saw that he had something in his hand that looked like money. Dr. Jackson heard defendant ask Dr. Brown what he had decided to do about the stuff, or words to that effect, and then Dr. Brown replied that he had not decided to do anything and asked defendant what he was talking about. Then Dr. Jackson went back to the kitchen and shortly thereafter heard a pistol shot. He went into the hallway and saw defendant with a pistol in his hand and saw him shoot a second time at Dr. Brown, who fell on the floor of the hallway. Defendant then stepped over the body and ran out of the house. Dr. Jackson pursued the defendant, but the latter turned and fired at him and escaped.

The Definitive Word

So what is the definitive word—Did James Alphonso Frye kill Dr. Robert Wade Brown? We shall defer to the collective wisdom of the jury who, according to U.S. Supreme Court Justice Clarence Thomas, " . . . determine the weight and credibility of witness testimony, . . . and are presumed to be fitted for it by their natural intelligence and their practical knowledge of men and the ways of men" (*U.S. v. Scheffer*, citing Aetna Life, 1891). Because the jury was convinced to a moral certainty of Frye's culpability, although not necessarily absolutely certain, they were right to convict. As we very well know, a conviction carries with it the connotation of guilt. Juries have nevertheless convicted innocent defendants in the past only to later learn that perhaps they were not as fit for their duty as Justice Thomas asserts. We do not believe this is one of those cases.

What we can state unequivocally is that the court found Frye guilty of murder in the second degree and presumably did so based on the evidence. With deliberations lasting less then one hour it is reasonable to assume that the jury's decision was not one that they agonized over.

Frye was the only person to ever confess to killing Dr. Brown. His confession was quite detailed and the only eyewitness to the killing, Dr. Julian Jackson, identified Frye as the killer and as the person who shot at him (Dr. Jackson) during his escape from Dr. Brown's residence.

Dr. Marston's discontinuous systolic blood pressure test results proved nothing. His test results were never entered into evidence even though the jury may have gained an insight into those test results during the trial.

There is no supporting evidence, other than Frye's own assertion of a conspiracy, which would lead a reasonable person to conclude that Frye was duped into confessing in order to secure half the reward money. Contrary to repeated assertions, there is no credible evidence to even suggest three years following his incarceration Frye was set free as a result of the "real killer" coming forward. No one else ever admitted killing Dr. Brown. Frye,

in fact, served a period of incarceration two months shy of 18 years. Finally, let us put to rest the myth that James Frye ever received Executive Clemency or Presidential pardon—he did not!

With so much misinformation in the literature and polygraph culture about James Frye, we hope to have debunked at least some of the myths. We also set out to explore the evidence and to independently answer definitively the question of Frye's culpability in the death of Dr. Brown. Despite Frye's intricate web of deceit we believe the evidence is sufficiently persuasive and speaks for itself. Unfortunately, we are resigned to the fact that the answers to all of the questions asked about the profession's most infamous examinee might never be known. Those answers are buried in grave Number 6230, Section 33, Arlington National Cemetery—James Alphonso Frye's final resting place.

Author Note

Three very old newspaper articles were used in preparation of this paper the source and date of which could not be identified precisely. Contextual clues suggested publication was between 1921 and 1922. The articles are entitled, "*Negro held on charge of slaying physician: Police say James Frye confessed to killing colored doctor last year*": "*Holds Frye guilty of killing doctor: Jury returns second-degree verdict after less than an hour's deliberation: Bars use of lie detector: Not perfected, Justice McCoy says, sufficiently to be introduced in court—big crowd present*"; and "*Dr. Brown Killed.*" These articles are available for review by contacting the authors.

Acknowledgements

The authors wish to acknowledge the work (1981) of Professor J. E. Starrs, The George Washington University, National Law Center, Washington, D.C. 20052. Professor Starrs presented a paper at the 33rd Annual Meeting of the American Academy of Forensic Sciences (February 17-20, 1981), Los Angeles, CA entitled: A "Still-Life Watercolor": *Frye v. United States*. Professor Starrs reviewed and analyzed the documentary material in *Frye v. United States* and also addressed the facts surrounding the investigation.

Chronology of Key Events in the Life of James Alphonso Frye

Apr	8	1895: Date of Birth (Report of Interment). Starrs (1983) Reports Other Documentary Evidence Listing Frye's Date of Birth as April 5, 1895
Aug	4	1918: Enlisted into U.S. Army
Jan	20	1919: Discharged from U.S. Army
Nov	27	1920: Dr. Robert W. Brown was shot and killed
Jul	21	1921: Frye and accomplices committed robbery of G.R. Blake
Aug	16	1921: Frye arrested by D.C. police for Robbery; Placed in Washington Asylum & Jail
Aug	18	1921: Frye arrested by USSS for Treasure Check Fraud
Aug	22	1921: Confessed to Murder of Dr. Brown
Feb	27	1922: Grand Jury Presentment, Murder in the First Degree

Mar	10	1922: Indictment for Murder in the First Degree
Mar	11	1922: Arraigned; Plead "Not Guilty"
Jul	14	1922: Motion for Continuance--Unavailability of Key Witness; Denied
Jul	17	1922: Trial Began; *United States vs. James Alphonso Frye*; Supreme Court of the District of Columbia, Criminal No. 38,325; Honorable Walter I. McCoy, Presiding Judge
Jul	20	1922: Date of Conviction; Verdict: Guilty, Murder in the Second Degree
Jul	28	1922: Date of Sentence; Supreme Court of District of Columbia Judgment and Noted Appeal Made in Open Court
Aug	3	1922: Application for Appeal
Aug	7	1922: Leave Granted to Prosecute Appeal in Forma Pauperis
Sep	26	1922: Bill of Exceptions Submitted
Nov	26	1922: Sentenced to Life Imprisonment
Dec	4	1922: Amendments to Bill of Exceptions Filed
Feb	8	1923: Assignment of Errors and Designation of Record Filed
Dec	3	1923: Judgment Affirmed by Court of Appeals of District of Columbia
Dec	21	1923: Mandate From Court of Appeals Affirming Judgment of Supreme Court of District of Columbia
Jan	5	1924: Warrant for Removal to Penitentiary
Feb	1	1924: Transferred to the U.S. Penitentiary, Fort Leavenworth, Kansas from Washington Asylum & Jail
Apr	13	1932: Transferred to D.C. Reformatory, Lorton, VA
July	12	1934: Application for Executive Clemency
July	21	1934: Letter from U.S. Attorney Leslie C. Garnett to The Attorney General in Response to Frye's Application for Executive Clemency
Apr	13	1932: Transfer from U.S. Penitentiary to Reformatory, Lorton, VA
Jul	21	1936: Application for Executive Clemency
Aug	1	1936: Acting Pardon Attorney Robert H. Turner Letter Rejecting Submission of Case to the President for Consideration
Jun	17	1939: Released on Parole by the Board of Indeterminate Sentences and Parole

Dec	20	1940: Letter From Assistant Pardon Attorney Turner to Board of Indeterminate Sentences and Parole, Denying Additional Consideration in Frye's Case by Way of Commutation of Sentence or Pardon
Aug	2	1943: Frye Letter to the Department of Justice, Pardon Attorney, Honorable Daniel M. Lyons
Sep	2	1943: Letter from Pardon Attorney Daniel M. Lyons to Frye Advising that He Serve a Longer Period of Probation Before Additional Case Review is Undertaken
Sep	7	1945: Letter to Pardon Attorney Regarding Petition for Executive Clemency
Sep	28	1945: Frye Letter to United States President Harry S. Truman
Oct	18	1945: Pardon Attorney's Response to Frye's Letters Regarding Procedures For Petition for Executive Clemency
Jan	28	1953: Date of Death
Feb	3	1953: Date of Interment

References

August, R. S. (1974). The polygraph and the courts. *Scientific Security.*

Bratton, B. (1921, August 3). Memorandum for the Chief, U.S. Secret Service. Treasury Department, United States Secret Service, Washington, DC.

Confesses to murder of colored physician: James A. Frye tells police shooting was done in self-defense after doctor struck him. (1921 August 23). *The Evening Star*, p. 5.

Daily legal record: Number 38,325. (1922 July 29). *The Washington Post*, p. 5.

Doctor Brown killed. {Article appeared in an unidentified and undated newspaper in 1921.}

Doctor killed over prescription: One charge—Say girl used mails wrongly. (1922 March 11). *The Washington Post*, p. 12.

Doctor shot dead; Assailant flees: Mystery surrounds killing of Dr. R. W. Brown in his home. (1920 November 28). *The Washington Post*, p. 2.

Daubert v. Merrill Dow Pharmaceuticals, Inc, 509, U.S. 579, 589 (1993).

Frye, J. (n.d.). *Report of Interment, grave number 6230, section 33.* Arlington National Cemetery.

Frye v. United States, 293 F. 1013 (D.C. Cir.1923).

Garnett, L. C. (1934, July 21). *Letter to the Pardon Attorney's Office, Department of Justice.* Office of the United States Attorney, District of Columbia, Washington, DC. *General Electric Co. et al. v. Joiner et ux.*, 522 U.S. 136 (1977).

Holds Frye guilty of killing doctor: Jury returns second-degree verdict after less than an hour's deliberation: Bars use of lie detector: Not perfected, justice McCoy says, sufficiently to be introduced in court—big crown present. {Article appeared in an unidentified and undated newspaper.}

Jasanoff, S.(1989). Science on the Witness Stand: Scientific evidence is interpreted differently in the court than in the lab, but experts' collective actions can be influential. *Issues in Science and Technology,* 6(Fall), pp.80-87

Kumho Tire Co., LTD., et al., v. Carmichael et al., 526 U.S. 137 (2000).

Larson, J. (1938). Introduction. In *The lie detector test.* American Polygraph Association, (pp. 1-3) NY: Richard Smith.

Lie Detector gets new chance in court: Appeal from life sentence noted by attorneys for negro James Frye. (1922, July 29). *The Evening Star*, p. 24.

Life sentence penalty in murder of doctor: Appeal taken for Frye, who was denied use of lie Detector—Yeoman jailed. (1922 July 29). *The Washington Post*, p. 8.

Marino, C. F. (1975). Labor and Legislation Panel. In N. Ansley, (Ed.), *Legal admissibility of the polygraph* (p. 325). Springfield, IL: Charles C. Thomas.

Marston, W. M. (1935, April). Can you beat the lie detector? *Esquire: 40*, (pp.174-177).

Marston, W. M. (n.d.) Lecture summary of can you beat the lie detector? *Celebrities for Educational and Entertainment Auspices of Coming Events*: {Advertisement Brochure}. New York: {Ernest Biggs}Times Building.

Marston, W. M. (1938). *The lie detector test.* American Polygraph Association, New York: Richard Smith, pp. 38-40.

Matte, J. A., (1980). *The art and science of the polygraph technique.* Springfield, IL: Charles Thomas, p. 9.

McCahill, M. (1921, August 16). *Special report to Mr. W. H. Moran, Chief, U.S. Secret Service.* Treasury Department, United States Secret Service, Washington, DC.

McCahill, M. (1921, August 23). *Special supplemental report to Mr. W. H. Moran, Chief U.S. Secret Service.* Treasury Department, United States Secret Service, Washington, DC.

McCahill, M. (1921, August 24). *Operative McCahill reports for Wednesday, August 24, 1921* {Daily report of agent}. Treasury Department, United States Secret Service, Washington, DC.

McCahill, M. (1922, July 28). *Operative McCahill reports for Friday, July 28,1922* {Daily report of agent}. Treasury Department, United States Secret Service, Washington, DC.

Mystery surrounds killing of Dr. R.W. Brown in his home (1920, November 28). *The Washington Post* (page unknown).

Negro held on charge of slaying physician: Police say James Frye confessed to killing colored doctor last year. {Article appeared in an unidentified and undated newspaper sometime in 1922}.

Pringle, H. F. (1936 September). How good is any lie? *The Reader's Digest:* Condensed from McCall's, 75-78.

Robinson, H. M. (1935, January). Science gets the confession, *The Reader's Digest: 26*, Condensed from Forum, 1-6.

Say he confessed killing: Police assert J.A. Frye, colored, admits shooting Dr. R.W. Brown. (1921 August 29). *The Washington Post*, p. 14.

Starrs, J. E. (1983). A still-life watercolor: Frye v. United States. *Polygraph, 12*, 76-90.

United States v. James Alphonso Frye. Amendments to the Bill of Exceptions. United States District Court for the District of Columbia Criminal Case File 38,325 (Record Group 21), National Archives, Suitland Reference Branch, Suitland, MD.

United States v. James Alphonso Frye. Application for Executive Clemency, National Archives, Suitland Reference Branch, (1936 July 21), Suitland, MD.

United States v. James Alphonso Frye. Request for Continuance. United States District Court for District of Columbia Criminal Case File 38,325 (Record Group 21), National Archives, Suitland Reference Branch (14 July1922), Suitland, MD.

United States v. James Alphonso Frye. United States Court of Appeals for District of Columbia Case File 3968 (Record Group 276), National Archives, Suitland Reference Branch, Suitland, MD.

United States v. James Alphonso Frye. United States District Court for District of Columbia Criminal Case File 38,325, Docket Entries. National Archives, Suitland Reference Branch, Suitland, MD.

United States v. James Alphonso Frye. United States District Court for District of Columbia Criminal Case File 38,325 (Record Group 21), National Archives, Suitland Reference Branch, Suitland, MD.

United States v. James Alphonso Frye. United States Pardon Attorney case file 56,386, National Archives, Suitland Reference Branch, Suitland, MD.

U.S. v. Scheffer, 523 U.S., 303 (1998) citing *Aetna Life Ins. Co. v. Ward*, 140, U.S. 76, 78 (1891).

Instructions to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript Submission

Manuscripts should be in English, and submitted, along with a cover letter, to Editor, American Polygraph Association, PO Box 10342, Ft. Jackson, South Carolina 29207 (USA). The cover letter should include a telephone number, return address, and e-mail address. Authors should also state clearly in the cover letter if they wish to submit their manuscript to a formal peer-review. The preferred method of manuscript submission is as an email attachment (MS Word, WordPerfect, or PDF format) with the cover letter included in the body of the email. Send to the Editor at:

StuandDee@msn.com

Authors without Internet access may also submit manuscripts on computer disk along with 5 paper copies to the editorial address above. As a condition for publication, authors shall be required to sign a statement that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere.

Manuscript Organization and Style

All manuscripts must be complete, balanced, and accurate. All authors should follow guidelines in the *Publication Manual of the American Psychological Association* (4th edition). The manual can be found in most public and university libraries, and can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psychophysiological Detection of Deception* included in this volume. Legal case citations should follow the *West* system.

Manuscript Review

A single Associate Editor will handle papers, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 12 weeks of receipt of manuscript. Articles submitted for publication are evaluated according to several criteria including significance of the contribution to the polygraph field, clarity, accuracy, and consistency.

Copy-editing

The Editor reserves the right to copy-edit manuscripts. All changes will be coordinated with the principal author.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be approved, if the senior author concurs.

Professional Copies

The senior author will receive ten (10) copies of the journal issue in which the article appears.