Gray’s three-arousal model: an empirical investigation

Peter A. Arnett\textsuperscript{a,}\textsuperscript{*}, Joseph P. Newman\textsuperscript{b}

\textsuperscript{a}Psychology Department, The Pennsylvania State University, University Park, PA 16802-3104, USA
\textsuperscript{b}Psychology Department, University of Wisconsin — Madison, Madison, WI 53706-1611, USA

Received 19 May 1998; received in revised form 7 June 1999; accepted 13 July 1999

Abstract

We evaluated the validity of Gray’s and Fowles’ three-arousal model in two studies of criminal offenders using a continuous motor task involving rewards and punishments. Consistent with predictions for the behavioral approach/activation system (BAS), offenders displayed significant ($p < 0.00001$) increases in response time and heart rate (HR) from a no-incentive practice phase to a reward-only (experiment 1) and active-avoidance (experiment 2) phase. Trait impulsivity was correlated with the response time index of BAS activation in experiment 1 but not experiment 2. Consistent with predictions for the behavioral inhibition system (BIS), offenders showed a significant increase in number of skin conductance responses (SCRs) ($p < 0.05$) from a reward-only to a mixed-incentive phase in experiment 1 and a significant increase in SC amplitude to punishment cues in both experiments. Consistent with predictions for the dynamics of the model, participants showed significant slowing of response time ($p < 0.0001$) from reward-only (experiment 1) or active avoidance (experiment 2) to mixed-incentive phases despite showing an initial tendency toward response facilitation to the onset of the punishment cue signifying the beginning of the mixed-incentive phases. Participants also showed significant ($p < 0.002$) decreases in HR between these phases in both studies, but this effect was only evident on trial 1. The BIS-influenced response time and HR indices were significantly ($p < 0.05$) correlated with anxiety in experiment 1, but unexpectedly, anxiety was not correlated with SC indices of the BIS in either study. Although much of the data support the validity of the Gray/Fowles model, particular findings suggest that further refinement of this theory is indicated. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Gray; Fowles; Three-arousal model; BIS; BAS; Psychopaths; Anxiety
In 1980, Fowles proposed a three-arousal model that was an adaptation and extension of Gray’s (1975) two-factor learning theory as an explanation of behavior in general and psychopathological behavior in particular. The three-arousal systems in this model are: (a) the behavioral approach (Gray) or activation (Fowles) system (BAS); (b) the behavioral inhibition system (BIS); and (c) a nonspecific arousal system (NAS) receiving inputs from both the BAS and the BIS. There are also mutually inhibitory inputs between the BAS and BIS. Incorporating the ideas of both Gray and Fowles, the BAS is activated by and initiates behavior in response to conditioned stimuli for reward or active avoidance and is indexed by heart rate (HR) and speed of responding. The BIS is activated by and inhibits behavior in response to conditioned stimuli for punishment (passive avoidance) or frustrative non-reward (extinction) and is indexed by increased electrodermal activity (EDA). Additionally, Newman and his colleagues have attempted to assess the dynamics of the three-arousal model in psychopaths (Arnett, Smith & Newman, 1997; Newman, Patterson, Howland & Nichols, 1990; Newman, Patterson & Kosson, 1987; Newman & Wallace, 1993) and extraverts (Nichols & Newman, 1986) by examining changes in response time to punishment cues after a dominant response set for reward has been established (see Newman (1997) for a detailed discussion of Gray’s model). They have hypothesized that, once a dominant response set for reward has been established in these disinhibited populations, the introduction of punishment cues should have the effect of increasing NAS activity and, consequently, accentuate ongoing behavior at the time (reward-seeking activity). Thus, rather than showing response inhibition following cues for punishment in these situations like other individuals, disinhibited individuals should show response facilitation (see Fig. 1).

Although Fowles (1980) did not emphasize this aspect of Gray’s writings, Newman and colleagues’ proposals regarding response facilitation follow directly from Gray’s (1987) claims regarding the consequences of nonspecific (i.e. general) arousal (see pp. 179–184 regarding the consequences of drive summation). Specifically, Gray’s model posits two consequences of BIS activation: one output from the BIS reduces the probability of ongoing approach behavior by shifting the motivational balance toward inhibition whereas the second output increases NAS arousal. Notably, however, if the BIS activation fails to inhibit ongoing approach behavior, then the increased NAS arousal associated with the BIS activation will serve to increase the vigor of the emitted response and result in response facilitation (i.e. drive summation) (see Fig. 1). Thus, an input to the BIS may paradoxically increase rather than decrease the intensity of approach behavior when situational or personality-related factors reduce the probability of behavioral inhibition (see Wallace, Bachorowski & Newman, 1991).

An important contribution of Fowles (1980) was his incorporation of autonomic psychophysiological indices into Gray’s (1975) model. Despite the frequency with which the Gray/Fowles model or variations of it have been used to explain syndromes of disinhibition such as psychopathy (e.g. Arnett, 1997; Arnett et al., 1997; Newman & Wallace, 1993) or syndromes of inhibition such as anxiety disorders (Geen, 1987; Hagopian & Ollendick, 1994), there has been little research examining the general validity of the constructs of the model. The current studies were designed to evaluate the validity of the model in a group of criminal offenders using a continuous motor task involving cues for reward, passive avoidance and active avoidance. Both autonomic psychophysiological and behavioral indices were measured throughout.
To our knowledge, only one previous study has examined the general validity of the Gray/Fowles model using self-report measures, behavioral indices and autonomic psychophysiological variables. Gomez and McLaren (1997) examined changes in heart rate (HR), skin conductance level (SCL) and mood in response to performance in either a reward or punishment condition of a go/no-go discrimination learning task. In the reward condition, participants were shown a set of 12, two-digit numbers presented sequentially on a computer monitor. Half of the numbers were “good” and the other half “bad,” and participants were instructed to distinguish the good from the bad numbers. They received a 10-cent monetary reward when they either responded to a good number or did not respond to a bad number. There was no loss of money for responding to a bad number or failing to respond to a good number. Participants in the punishment condition lost 10 cents for either responding to a bad number or failing to respond to a good number. Consistent with the Gray/Fowles model, compared with participants in the punishment condition, participants in the reward condition made more commission errors, reported significantly higher happy mood scores and significantly lower nervous mood scores, and had lower SCL. These findings were consistent with these authors’ predictions from the model, but others were not. For instance, they predicted that HR would increase more in the reward condition than in the punishment

Fig. 1. Gray’s (1987) Neuropsychological model of approach-avoidance learning. The reward system in the figure is the behavioral activation system (BAS) and is activated by both reward and active avoidance cues; the punishment system is the behavioral inhibition system (BIS) and is activated by punishment cues. From The Psychology of Fear and Stress by J. A. Gray, 1987, New York: Cambridge University Press. Copyright 1987 by Cambridge University Press. Adapted by Permission.
condition, but it did not. Other findings inconsistent with these investigators’ hypotheses from the Gray/Fowles model were that HR was not associated with happy mood ratings, any subscales of Eysenck’s Impulsiveness Scale (Eysenck & Eysenck, 1977) or any of Carver and White’s (1994) BAS scales. Additionally, SCL did not correlate with Spielberger’s STAI-trait form (Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983) or Carver and White’s (1994) BIS scale.

There are at least four possible reasons why Gomez and McLaren (1997) failed to find the predicted association between HR and reward. First, in contrast to Fowles’ (1980) theorizing, HR may simply not be an accurate indicator of BAS activity. Second, it may be that, although participants did not receive punishment for responding to bad numbers or failing to respond to good numbers in the reward condition, their failure to obtain the monetary reward on these trials resulted in some BIS activation that dampened any ongoing BAS activity. In Gray’s (1987) model, the BIS is activated by ‘frustrative nonreward’ as well as punishment cues. On the trials where participants responded to a bad number or failed to respond to a good number, they were presumably expecting to obtain reward. Third, it may be that for HR to be an adequate index of BAS activity and increase significantly to reward cues, an approach response with a more demanding motor component than the one used in Gomez and McLaren’s (1997) study is required. Finally, these investigators may not have found significant differences between their reward and punishment conditions because their punishment condition involved a passive and active avoidance response. According to Gray, active avoidance is mediated by the BAS and thus, as Fowles (1980) theorizes, should generate increased HR similar to reward responding. Consequently, the BAS would be activated in both conditions of Gomez and McLaren’s (1997) study and may have limited the opportunity to observe significant differences in HR.

Gomez and McLaren’s (1997) failure to find significant correlations between their SCL index in the punishment condition and anxiety is surprising. However, it is possible that significant correlations would have emerged if these investigators had taken a different data analytic strategy. As Fowles (1980) notes, phasic skin conductance responses (SCRs) to a specific punishment cue and spontaneous fluctuations in SCRs in situations where punishment is anticipated are more reliable indicators of anxiety or BIS activity than tonic SCL changes. Gomez and McLaren’s (1997) use of tonic changes in SCL may have limited their ability to find significant correlations with anxiety measures because it is not a particularly sensitive measure of BIS activation.

In the current experiments, we attempted to improve upon Gomez and McLaren’s (1997) study by developing a task that minimized the possible effects of frustrative non-reward on BAS activation and involved a much more demanding motor component. We also used spontaneous fluctuations in SC and SC amplitude to punishment cues because they have been shown to be more sensitive indices of BIS activation than tonic SCL changes. We attempted to extend Gomez and McLaren’s (1997) findings by examining behavioral and psychophysiological responses to active avoidance cues in addition to reward cues as elicitors of the BAS. Additionally, in both of our experiments we examined all elements of the Gray/Fowles model within the same groups of participants rather than examining one aspect (e.g. BAS) in one group and another aspect (e.g. BIS) in another group. Finally, we attempted to examine the dynamics of the Gray/Fowles model.
It is important to note that the data presented for the studies outlined here were gathered in the same experiments and with the same subjects as for our recently published paper (Arnett et al., 1997). This previous paper involved evaluating three different theories of psychopathic behavior that relied on constructs from Gray's model. Although there is some overlap with this paper, the focus of the current paper is on evaluating the generality of Fowles' adaptation of Gray's model and its relationship to broad personality dimensions of anxiety and impulsivity.

1. Method

1.1. Participants

There were 58 participants in experiment 1 and 64 participants in experiment 2. All participants were Caucasian prison inmates at a minimum security prison in southern Wisconsin. After determining each inmate's eligibility for participation by briefly reviewing his institutional file, potential participants were selected by identifying every fifth name on the institution roster. Inmates were excluded from eligibility for participation if they were: (1) above age 40 or below age 18, (2) below the fourth grade level on standardized achievement tests or (3) identified as actively psychotic or taking psychotropic medication. Additionally, only those participants receiving a diagnosis of 'psychopath' or 'non-psychopath' based on standard cutoffs from Hare's (1985) Psychopathy Checklist were included (see Arnett et al. (1997) for more details of this selection procedure).

1.2. Task and apparatus and measures

For both studies, HR and EDA were recorded while participants performed a modified version of a continuous motor task employed by Fowles, Fisher and Tranel (1982) and modified and described in further detail by Arnett et al. (1997). A schematic of the task is also available in Arnett, Fischer and Newby (1996).

The apparatus was identical for both experiments and consisted of five buttons arranged on a board in a semi-circle. A red and a green light were mounted above each button. In addition, one button was mounted at the center of the board with a larger green and red light above it. Each press of the center button activated one of the outer green lights indicating that participants should press that button as quickly as possible. When the center red light was on, an outer green light would occasionally switch to red. In the reward experiment (experiment 1), participants won five cents after every five responses if their response time exceeded criterion but lost 25 cents if they failed to inhibit responding when a red light appeared. In the active avoidance experiment (experiment 2), participants lost five cents if their response time did not exceed criterion, and like experiment 1, lost 25 cents for failing to inhibit responses to red lights. Passive avoidance errors were operationalized as failures to inhibit responding to these red lights in both experiments. The response time criterion was continually adjusted by the computer controlling the task, so that all participants won five cents on approximately 80% of their five response sequences during the reward-only phase (experiment 1) or avoided losing
five cents on approximately 80% of their five response sequences during the active avoidance phase (experiment 2).

There were four, 2-min trials in the task. The first minute of each trial was a reward-only phase (experiment 1) or an active avoidance phase (experiment 2) signaled by the center green light onset. For both experiments, the second minute involved the possibility of punishment and was signaled by the center red light onset (mixed-incentive phase). The two phases were continuous, that is, there was no pause or break between them. The four trials were separated by 1-min rest phases and were preceded by a 2-min resting baseline and a 1-min practice phase during which participants made button presses to green lights without incentives. Poker chips worth five cents were used for monetary incentives.

Participants also completed the Welsh Anxiety Scale (1956). We chose this particular measure of anxiety because, according to Gray (1991), the Welsh Anxiety Scale corresponds to his conceptualization of anxiety consisting of two parts neuroticism and one part introversion. We also included an impulsivity scale derived from MMPI items described by Blackburn (1971). This measure was chosen because it provides a reasonable operationalization of Gray’s (1970,1987) conceptualization of impulsivity in that Blackburn’s measure has been shown to be significantly positively correlated with both Neuroticism and Extraversion as measured by the Eysenck Personality Inventory. Gray views impulsivity as representing an interaction between Eysenck’s (1967) Extraversion–Introversion and Neuroticism–Stability dimensions, with high impulsivity associated with the neurotic extravert quadrant. Finally, to obtain an estimate of Wechsler Adult Intelligence Scale — Revised (WAIS-R) Full Scale IQ, participants were given the Shipley Institute of Living Scale (Zachary, 1986).

1.3. Psychophysiological recording

Data processing and storage were controlled by a Zenith (model ZF-158-41) computer. The following physiological signals were processed using a Beckman R511A dynograph and digitized using a Scientific Solutions, 12 bit DAS, 40 kHz A/D board.

1.3.1. Heart rate

ECG was recorded by attaching an electrode to a rib on the right and left side of the torso using adhesive collars and Beckman Standard 1 cm² Ag–AgCl electrodes with Spectra 360 electrode gel as the conducting medium. Prior to electrode placement, the participant’s skin was abraded using gauze moistened with rubbing alcohol. The ECG was recorded using a Beckman Type 9806A AC coupler with a 0.1-s time constant. The output was directly digitized and recorded by the computer at a rate of 100 Hz.

1.3.2. Skin conductance

Skin conductance was recorded from the second and third digits of the non-dominant hand using Beckman Standard 1 cm² Ag–AgCl electrodes. A Unibase and saline mixture was used as the conducting medium (see Fowles et al., 1981, p. 235 for formula). Skin conductance signals were recorded through a constant-voltage Lykken skin conductance coupler. The output was directly digitized and recorded by the computer at a rate of 20 Hz.
1.4. Procedure

All participants meeting the selection criteria were contacted about participating in a study involving an initial one-and-a-half-hour interview and several behavioral tasks providing the opportunity to earn money. Participants were paid US$3.00 for the interview and recontacted within 1–4 weeks for the experiment described here. The interview data were not used in the current investigation and are described in more detail in Arnett et al. (1997).

Prior to the experiment, all participants were naive to the experimental situation; they did not participate in any previous experiments. Additionally, separate groups of participants were run for each experiment, such that no individual participated in both experiments. Each participant was run on the task by one of two male experimenters blind to participants’ psychopathy, anxiety, or impulsivity status. After signing a consent form on the testing day, color-blindness was assessed using Ishihara Tests for Colour Blindness (Ishihara, 1989). If the participant had normal color vision, he was asked to wash his hands as a prelude to attaching SC electrodes. The participant then completed the questionnaire containing the Welsh Anxiety Scale and Blackburn’s impulsivity scale in a separate waiting room. After a minimum of 20 min and a maximum of 30 min, the participant returned to the experimental room, the ECG electrodes were attached, and a photo-plethysmograph was attached to his non-dominant thumb. After making the necessary adjustments on the psychophysiological recording equipment, the participant was told that a 2-min baseline of his physiology would be taken.

Following the baseline, the experimenter began reading the task instructions. After basic instructions for pressing buttons to lights were given, the participant began the 1-min practice phase. Following this, the experimenter told the participant for the first time that he would be playing the task for money. At this point the experimenter explained the monetary contingencies to the participant using a standard set of instructions, and poker chips were laid out on a table next to where the participant sat during the task. Finally, the participant was reminded again to try and win as much money as possible and asked if he had any questions prior to starting the task. In experiment 1, all participants began the task with US$1.00 or 20 chips; in experiment 2, they began the task with US$7.00, or 140 chips. The experimenter informed the participant of his winnings at the end of each rest period.

At the end of the task, the participant was told that his winnings would be deposited in his institution account. Finally, participants completed the Shipley and were thanked for participating.

1.5. Data reduction

1.5.1. Heart rate

The off-line ECG was edited with a computer program allowing visual inspection of the ECG array to identify and omit invalid heart periods. The remaining heart periods were then converted to second-by-second HR (Graham, 1978).

1.5.2. Skin conductance

Skin conductance responses were identified from the digitized data by a Pascal
implementation of the WAVE SC scoring program developed by Strayer and Macias (1982). Responses greater than or equal to 0.05 mS were identified.

1.6. Operationalization of constructs and predictions

The constructs from the Gray/Fowles model were operationalized using behavioral, cardiovascular and electrodermal indices. Activation of the BAS was indexed by HR and response time increase from the no-incentive practice phase to the reward (experiment 1) or active avoidance (experiment 2) phases. Activation of the BIS was indexed by increased EDA from the reward (experiment 1) and active avoidance (experiment 2) phases to the mixed-incentive phases, as well as by SC amplitude to the most salient punishment cues in both studies. The dynamics of the model were evaluated by examining response time change in the first 5 s following the most salient punishment cue — the center red light indicating the beginning of the mixed-incentive phase. Response time and HR slowing from the reward (experiment 1) and active avoidance (experiment 2) phases to the mixed-incentive phases were also used to measure the dynamics of the model.

For the BAS, it was predicted from the Gray/Fowles model that participants would show significant increases in HR and response time from the no-incentive practice phase to the reward-only phases (experiment 1) and active-avoidance phases (experiment 2). For the BIS, significant increases in SCRs from the reward-only (experiment 1) or active-avoidance (experiment 2) to the mixed-incentive phases were predicted. Additionally, significant increase in SC amplitude to the center red light onsets was predicted for both studies. For the dynamics of the model, response time and HR slowing were predicted to occur in the face of BIS cues (mixed-incentive phases) because of the inhibitory input from the BIS to the BAS (see Fig. 1). Response time and HR are measures of BAS activation, but the extent to which participants show a reduction in these indices when BIS cues appear is conceptualized as reflecting the degree of BIS inhibition of the BAS.

To explore the dynamics of Gray’s model further, we also examined participants’ reaction to the center red light, which is the most salient punishment cue used in the task. However, it is difficult to offer a simple hypothesis based on Gray’s model. Owing to the combined influences of the NAS and the decision mechanism which responds to the motivational balance of BAS and BIS inputs, Gray (1987, p. 246) writes “as threats of punishment for approach behaviour are increased, the probability of approach behaviour goes down, while the vigour of approach behaviour, if it nevertheless occurs, increases”. In the current task, then, the center red light might either (a) increase behavioral inhibition to the extent that it overcomes behavioral approach and results in slowing down or (b) produce behavioral facilitation (i.e. speeding up) if the output from the BIS to the NAS serves to increase arousal and the output from the BIS is insufficient to interrupt BAS-mediated approach behavior. To examine this reaction, we focused on the first 5 s after the punishment cue onset to gauge participants’ immediate reaction to the red light. In doing so, we believe that we increased the likelihood of observing response facilitation because the BAS influences (i.e. approach and active avoidance motivation) from the reward-only (experiment 1) or active avoidance phases will still be strong and the NAS arousal associated with the onset of the center red light will be most powerful. As the mixed-incentive phase progresses, we believe that it is increasingly likely that NAS
arousal associated with the onset of the red light will dissipate and that BIS-related awareness of the increased threat of punishment will overshadow BAS-related influences, making behavioral inhibition the more likely response.

Because the BIS is thought to underlie anxiety, and the BAS is thought to underlie impulsivity (Gray, 1990), BIS and BAS indices were correlated with the measures of anxiety and impulsivity, respectively. Larger increases in HR and response time from the no-incentive practice phase to the reward-only phase (experiment 1) or active avoidance phase (experiment 2) were predicted to be associated with higher self-reported impulsivity scores. Larger increases in number of SCRs and greater increase in SC amplitude to center red light onsets were predicted to be associated with higher self-reported anxiety scores. Additionally, because they were conceptualized to reflect the degree of BIS dampening of BAS activation, decreases in HR and response time from the reward-only (experiment 1) or active avoidance (experiment 2) to the mixed-incentive phases were also predicted to be associated with self-reported anxiety. Finally, response time change in the first 5 s following the onset of the mixed-incentive phases in both experiments were expected to be correlated with anxiety.

Indices measuring the BAS were intercorrelated with one another as were BIS-influenced indices to evaluate the extent to which they were measuring the same construct. Finally, number of passive avoidance errors were correlated with all personality, behavioral, and psychophysiological measures to evaluate the extent to which BIS, BAS, and model dynamic indices were related to task performance. For all correlational analyses, difference scores using the repeated measures values outlined below were calculated for the relevant variables.

2. Results

2.1. Data analytic strategy

To evaluate response time, HR and EDA changes from one part of the task to another, data were collapsed within phases and repeated measures analyses of variance (ANOVA’s) were conducted. To assess BAS activation using HR (bpm) and response time (ms) changes, median no-incentive practice phase values (both experiments) and the mean of the median reward-only phase values (experiment 1) or median active avoidance phase values (experiment 2) were used as repeated measures. To assess BIS activation using number of SCRs, repeated measures were mean reward-only (experiment 1) or active avoidance phase (experiment 2) values and mixed-incentive phase values. To assess BIS activation using SC amplitude, repeated measures were baseline SC level (set to ‘0’ for the purposes of these analyses) and mean amplitude (in mS) of SC response to center red light onsets. Our baseline SC level was admittedly artificially created, but selected as the best way to conduct a change analysis with phasic SC amplitude values because our SC program did not record SC level, only phasic SC changes. To assess the dynamics of the model using response time (ms) and HR (bpm), the mean of the median reward-only phase values (experiment 1) or median active avoidance phase values (experiment 2) and median mixed-incentive phase values were used as repeated measures. Additionally, mean response time in the last 10 s of the reward (experiment 1) or active avoidance (experiment 2) phases and mean response time in the first 5 s of the mixed-incentive phase were
used as repeated measures. We also conducted all analyses for all dependent variables on the first trial only, to evaluate the possibility that the findings may have been different when task contingencies were novel. Only in the case of HR decrease to BIS cues in both studies were the results substantially different than when the means collapsed across all four trials were used. These exceptions are discussed below.

Table 1 shows the participant characteristics of the sample.

2.2. Hypothesis testing analyses, experiment 1

2.2.1. BAS

Consistent with predictions, participants showed significant increases in response time \((F(1, 57)=107.7, p < 0.00001)\) and HR \((F(1, 57)=48.64, p < 0.00001)\) from the no-incentive practice to the reward-only phases (see Table 2). The impulsivity scale was significantly correlated with response time increases from practice to reward-only phases \(r(58)=0.27, p < 0.05\), but not HR \(r(58)=0.14, \text{ns}\).

2.2.2. BIS

Consistent with hypotheses, offenders showed a significant increase in SC amplitude to the center red light onsets, \(F(1, 57)=50.90, p < 0.0001\). Also consistent with expectations, participants showed a significant increase in the number of SCR’s from the reward-only to the mixed-incentive phases, \(F(1, 57)=5.67, p < 0.05\) (see Table 2). Contrary to expectations, neither change in SCR’s from the reward-only to mixed-incentive phases nor SC amplitude increase to center red light onsets were significantly correlated with anxiety \(r(58)=0.05, \text{ns}\) and \(r(58)=0.03, \text{ns}\), respectively).

2.2.3. Model dynamics

Consistent with predictions, participants showed significant slowing of response time from the reward-only to mixed-incentive phases \(F(1, 57)=139.41, p < 0.00001\). Contrary to predictions, participants did not show a significant decrease in HR from the reward-only to the mixed-incentive phases, \(F(1, 57)=3.06, p < 0.10\). However, when participants’ performance on the first trial was examined, the predicted decrease in HR occurred, \(F(1, 57)=11.69, p < 0.002\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Experiment 1, mean (S.D.)</th>
<th>Experiment 2, mean (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>58(^b)</td>
<td>64</td>
</tr>
<tr>
<td>Age</td>
<td>27.6 (5.4)</td>
<td>27.9 (6.1)</td>
</tr>
<tr>
<td>IQ</td>
<td>97.1 (10.3)</td>
<td>99.5 (11.8)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>10.2 (8.3)</td>
<td>12.2 (8.9)</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>16.3 (6.4)</td>
<td>17.3 (5.6)</td>
</tr>
</tbody>
</table>

\(^a\)S.D. means standard deviation, IQ Shipley Institute of Living Scale WAIS-R IQ estimate, Anxiety Welsh’s (1956) Anxiety Scale, Impulsivity Blackburn’s (1971) impulsivity scale.

\(^b\)Except for IQ, where \(n = 56\).
Supporting predictions, higher anxiety scores predicted greater HR decrease from the reward-only to mixed-incentive phases \(r(58) = 0.23, p < 0.05\) and greater response time slowing \(r(58) = 0.26, p < 0.05\). Consistent with the hypothesis that a punishment input following BAS activation results in response facilitation, participants showed a significant increase in response time following the onset of the most salient punishment cue, \(F(1, 57) = 9.23, p < 0.005\). This response facilitation measure was significantly correlated with impulsivity \(r(58) = 0.25, p < 0.05\), but not anxiety \(r(58) = -0.10, \text{ns}\).

### 2.2.4. Relationship between BAS indices

Increase in response time and HR from the practice to reward-only phases were significantly correlated \(r (58) = 0.46, p < 0.0001\).

### 2.2.5. Relationship between BIS and BIS-influenced indices

Change in number of SCRs from the reward-only to mixed-incentive phases was not significantly correlated with SC amplitude to the most salient punishment cues \(r(58) = 0.12, \text{ns}\), changes in response time \(r(58) = -0.16, \text{ns}\) or HR \(r(58) = 0.08, \text{ns}\). However, decreases in HR between these phases was significantly correlated with response time slowing \(r(58) = 0.34, p < 0.01\) and SC amplitude to the punishment cues \(r(58) = 0.28, p < 0.05\). This latter index was not significantly correlated with response time slowing \(r(58) = 0.01, \text{ns}\).

### Table 2

Behavioral and psychophysiological data, experiments 1 and 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Experiment 1, mean (S.D.)</th>
<th>Experiment 2, mean (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral and cardiovascular indices of BAS activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response time (ms), practice</td>
<td>1146.5 (223.4)</td>
<td>1234.1 (361.9)</td>
</tr>
<tr>
<td>Response time (ms), BAS phase</td>
<td>912.4 (136.5)</td>
<td>905.1 (160.5)</td>
</tr>
<tr>
<td>Heart rate (bpm), practice</td>
<td>83.6 (11.5)</td>
<td>81.1 (10.5)</td>
</tr>
<tr>
<td>Heart rate (bpm), BAS phase</td>
<td>89.5 (13.5)</td>
<td>87.0 (10.8)</td>
</tr>
<tr>
<td>Electrodermal indices of BIS activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR amplitude (mS), Center Red</td>
<td>0.23 (0.24)</td>
<td>0.17 (0.16)</td>
</tr>
<tr>
<td>SCR number, BAS phase</td>
<td>5.55 (2.49)</td>
<td>6.42 (2.36)</td>
</tr>
<tr>
<td>SCR number, mixed-incentive</td>
<td>5.99 (2.30)</td>
<td>6.22 (2.06)</td>
</tr>
<tr>
<td>Behavioral and cardiovascular indices of model dynamics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response time (ms), BAS phase</td>
<td>912.4 (136.5)</td>
<td>905.1 (160.5)</td>
</tr>
<tr>
<td>Response time (ms), mixed-incentive</td>
<td>999.8 (157.3)</td>
<td>962.6 (152.0)</td>
</tr>
<tr>
<td>Heart rate (bpm), BAS phase, 1st trial</td>
<td>93.2 (13.5)</td>
<td>88.8 (11.6)</td>
</tr>
<tr>
<td>Heart rate (bpm), mixed-incentive, 1st trial</td>
<td>90.0 (16.2)</td>
<td>86.5 (11.3)</td>
</tr>
<tr>
<td>Response time (ms), last 10 s, BAS phase</td>
<td>1017.2 (146.6)</td>
<td>1012.7 (175.1)</td>
</tr>
<tr>
<td>Response time (ms), first 5 s, mixed</td>
<td>981.6 (147.5)</td>
<td>933.7 (159.4)</td>
</tr>
</tbody>
</table>

\(^a\) S.D. means standard deviation, SCR skin conductance response, ms milliseconds, bpm beats per min, BIS behavioral inhibition system, BAS behavioral activation system. BAS phase is the ‘reward-only’ phase for experiment 1 and the ‘active-avoidance’ phase for experiment 2. Values for reward-only and active avoidance phases represent the means of four trials of median values for response time and heart rate and the mean of the four trials for skin conductance measures. Value for practice phase represents the median for response time and for heart rate.
2.2.6. Relationship between task performance and personality, behavioral and psychophysiological variables

Total passive avoidance errors on the task were significantly \((p < 0.05)\) correlated only with HR increase to reward cues \(r (58)=0.25, p < 0.05\), HR slowing between phases during the first trial of the task \(r (58)=0.24, p < 0.05\), response time slowing between phases overall \(r (58)=0.34, p < 0.005\) and response time change in the first 5 s following the onset of the center red light \(r (58)=0.35, p < 0.01\). Thus, participants who made more errors on the task showed a greater response time increase to reward cues, greater HR and response time slowing from the reward-only to mixed-incentive phases and greater response facilitation initially to the most salient punishment cue onset.

2.3. Hypothesis testing analyses, experiment 2

2.3.1. BAS

Consistent with predictions, offenders showed significant increases in response time \((F (1, 63)=114.5, p < 0.00001)\) and HR \((F (1, 63)=116.2, p < 0.00001)\) from the no-incentive practice to the active-avoidance phases (see Table 2). However, contrary to predictions, the impulsivity scale was not significantly correlated with response time or HR increases from practice to active-avoidance phases \(r (64)=0.01, ns, r (64)=0.13, ns\), respectively).

2.3.2. BIS

Consistent with predictions, participants showed a significant increase in SC amplitude to the punishment cues, \(F (1, 64)=67.96, p < 0.0001\); however, contrary to hypotheses, there was no significant change in number of SCRs from the active-avoidance to mixed-incentive phases, \(F (1, 63)=1.59, ns\) (see Table 2). Unexpectedly, the anxiety scale was not significantly correlated with change in number of SCRs \(r (64)=-0.05, ns\), or SC amplitude to the punishment cues \(r (64)=0.16, ns\).

2.3.3. Model dynamics

Consistent with expectations, participants showed significant slowing of response time from the active-avoidance to mixed-incentive phases \((F (1, 63)=80.0, p < 0.00001)\). Participants did not show a significant decrease in HR from the active-avoidance to the mixed-incentive phases, \(F (1, 63)<1.0, ns\); however, similar to experiment 1, when their performance on the first trial was examined, the predicted decrease in HR occurred, \(F (1, 63)=16.99, p < 0.0001\) (see Table 2). The anxiety scale was not significantly correlated with HR decrease from the active-avoidance to mixed-incentive phases \(r (64)=0.13, ns\) or response time slowing \(r (64)=0.14, ns\). Consistent with the hypothesis that a punishment input following BAS activation results in response facilitation, participants showed a significant increase in response time following the onset of the most salient punishment cue, \(F (1, 63)=15.70, p < 0.001\) (see Table 2). Similar to experiment 1, this response facilitation measure was significantly correlated with impulsivity \(r (64)=0.24, p < 0.05\), but not anxiety \(r (64)=0.19, ns\).
2.3.4. Relationship between BAS indices

Increase in response time and HR from the practice to active-avoidance phases were uncorrelated, $r(64)=0.02$, ns.

2.3.5. Relationship between BIS and BIS-influenced indices

Change in number of SCRs from the active-avoidance to mixed-incentive phases was significantly correlated with SC amplitude to the punishment cues $r(64)=0.22$, $p < 0.05$, but not with changes in response time $r(64)=-0.05$, ns) or HR $r(64)=0.03$, ns). Additionally, SC amplitude to the punishment cues was not significantly correlated with HR $r(64)=-0.07$, ns) or response time $r(1, 64)=-0.08$, ns) slowing. Finally, decrease in HR between phases was not significantly correlated with response time slowing, $r(64)=0.04$, ns).

2.3.6. Relationship between task performance and personality, behavioral, and psychophysiological variables

Total passive avoidance errors on the task were significantly ($p < 0.05$) correlated with HR slowing between phases during the task overall $r(64)=0.40$, $p < 0.005$), and negatively correlated with response time speeding up from the practice phase to the active avoidance phases overall $r(64)=-0.23$, $p < 0.05$). There were also marginally significant effects ($p < 0.10$) for response time slowing between phases overall $r(64)=0.20$, $p < 0.06$), and HR slowing between phases on the first trial $r(64)=0.20$, $p < 0.06$). Thus, similar to experiment 1, participants who made more errors on the task showed greater HR and response time slowing between phases. In contrast to experiment 1, participants who sped up more during the first (active avoidance) part of the task made fewer errors on the task.

2.4. Follow-up analyses

One particularly interesting result was our finding of a significant increase in number of SCRs from the reward to the mixed-incentive phase in experiment 1 in contrast to our null finding when number of SCRs from the active avoidance to the mixed-incentive phase was examined in experiment 2. According to Fowles’ (1980) formulation, active avoidance and reward cues activate the BAS. Therefore, increases in SCRs (BIS index) from reward to BIS activation and from active avoidance to BIS activation should be comparable. However, in examining our electrodermal data, it appeared that the BIS was activated similarly during both the active avoidance and mixed-incentive phases in experiment 2, at least after participants’ initial response to the salient punishment cue indicating the onset of the second phase of the task. This led us to the (post-hoc) interpretation that, in addition to activating the BAS, active avoidance cues might activate the BIS because they are, in fact, cues for punishment. Therefore, we conducted a follow-up analysis using number of SCRs in the BAS (reward or active avoidance) phase and the BIS (mixed-incentive) phase as the within subjects factors and condition (experiment 1 or 2) as the between-subjects factor. There was a significant condition $\times$ phase interaction, $F(1, 119)=6.85$, $p = 0.01$. Examination of Table 2 reveals that this effect is due to the relatively greater number of SCRs shown by participants in the active avoidance phase compared with the reward-only phase.

Surprisingly, neither of our primary indices of the BIS — SC amplitude to the punishment
cues and change in number of SCRs from the reward/active avoidance phases to the mixed-incentive phases — were correlated with anxiety in either study. Because Newman, Wallace, Schmitt and Arnett (1997) have suggested that anxiety may operate differently in nonpsychopaths compared with psychopaths, we reanalyzed our SC/Welsh Anxiety Scale correlations using only the nonpsychopaths in both studies. None of these follow-up correlations were significant ($r$'s ranging from —0.07 to 0.23, all $p$'s > 0.10), suggesting that the inclusion of psychopathic individuals in our study was not responsible for the nonsignificant correlations between our primary BIS indices and our anxiety measure.

3. Discussion

The current investigation incorporated behavioral, self-report and autonomic psychophysiological indices to evaluate the validity of the Gray/Fowles three-arousal model. Consistent with predictions for the BAS, participants demonstrated significant increases in response time and HR when the BAS was activated by reward (experiment 1) or active avoidance (experiment 2) cues. Additionally, the response time index of the BAS in experiment 1 was significantly correlated with the self-report impulsivity measure. Finally, response time and HR indices of the BAS were significantly correlated in experiment 1. Contrary to expectations, the HR index of the BAS in experiment 1, and both HR and response time indices of the BAS in experiment 2 were not significantly correlated with the impulsivity measure. Additionally, response time and HR indices of the BAS were uncorrelated in experiment 2.

Consistent with hypotheses for the BIS, participants showed significant increases in SCR amplitude to punishment cues in both experiments, and also a significant increase in number of SCRs from the reward-only to mixed-incentive phases in experiment 1. Also consistent with expectations, change in number of SCRs from the active-avoidance to mixed-incentive phases was significantly correlated with SCR amplitude to the most salient punishment cue in experiment 2. Contrary to expectations, no significant increase in number of SCRs occurred from the active-avoidance to mixed-incentive phases in experiment 2, and change in number of SCRs from the reward-only to mixed-incentive phases was not significantly correlated with SC amplitude to the most salient punishment cues in experiment 1. Also surprisingly, the anxiety measure was not correlated with either SCR measure in either experiment.

Consistent with hypotheses regarding the dynamics of the Gray/Fowles model that the introduction of BIS cues should dampen ongoing BAS activation, participants showed significant decreases in HR and response time from the reward-only (experiment 1) or active-avoidance (experiment 2) phases to the mixed-incentive phases. In addition, these BIS-influenced HR and response time indices were significantly correlated with self-reported anxiety in experiment 1. Also, these BIS-influenced indices were correlated with each other in experiment 1, and the HR index was significantly correlated with one of the BIS indices, SC amplitude to the punishment cues. In contrast to expectations, this latter BIS index was not significantly correlated with the BIS-influenced response time index. Additionally, change in number of SCRs between phases was not correlated with the BIS-influenced HR or response time indices in either experiment, and SC amplitude to punishment cues was not correlated
with these indices in experiment 2. Also unexpectedly, the self-report anxiety measure was not correlated with BIS-influenced HR or response time in experiment 2. Finally, the BIS-influenced HR and response time indices were not correlated in experiment 2.

We explored the dynamics of the model in another way by examining participants’ response time change to punishment cues immediately following the establishment of a dominant BAS set. Our results suggest that, at least in this sample of criminal offenders, the modal initial behavioral response to punishment cues following the establishment of a dominant BAS-mediated response set was facilitation. This suggests that, rather than a punishment cue input in this situation inhibiting ongoing BAS-mediated behavioral responding, it initially results in accentuating this ongoing dominant response, mediated by the BIS influence on the NAS. Clearly, all participants eventually began slowing down their response rate when the punishment cues appeared, as reflected in their overall slower response time during mixed-incentive compared with either reward-only or active avoidance phases. Nonetheless, their initial reaction was to speed up relative to their response time just prior to the punishment cue onset, a response that may have been mediated by the BIS influence on the NAS following the punishment cue input. As noted earlier, in the event that the BIS fails to inhibit ongoing approach behavior, Gray’s model holds that the increased NAS activity will facilitate that behavior (see Fig. 1).

One could argue that our response facilitation results were more a function of participants slowing down at the end of the reward-only (experiment 1) or active-avoidance (experiment 2) phases than speeding up in response to the punishment cue at the onset of the mixed-incentive phases. As a comparison of the response time values in Table 2 illustrates, participants’ response time in the last 10 s of the reward-only and active avoidance phases was over 100 ms slower than the mean of their median response time over the entire reward-only and active avoidance phases, respectively. In contrast, their response time in the first 5 s of the mixed-incentive phases was only slightly faster (about 18 ms for experiment 1, and 29 ms for experiment 2) than the mean of their median response time during these phases overall. We believe that the most parsimonious interpretation of participants slowing down at the end of the reward-only/active avoidance phases is that it represents a trend toward fatigue at the end of these phases after participants have been responding continuously for 50 s. However, fatigue effects alone cannot account for participants’ subsequent response facilitation after this initial slowing. Because there was no pause between phases, this trend toward fatigue and slowed responding would have continued from s 61–65 (first 5 s of the mixed-incentive phase), but it did not. There is no reason to believe that subjects would suddenly begin responding significantly faster again at this point unless something significant happened. There is only one thing that happens between the last 10 s of the reward-only/active avoidance phases and the first 5 s of the mixed-incentive phases: the onset of the center red light, a very salient BIS cue. The fact that our response facilitation index correlated with the impulsivity dimension in both studies provides further validation to our view that the response facilitation occurring in the first 5 s of the mixed-incentive phases represents something psychologically meaningful in the context of Gray’s model, and is not simply a fatigue effect and function of the slowing that occurs at the end of the reward-only/active avoidance phases.

The results of our investigation suggest that, although many predictions from the Gray/Fowles model were supported, there appear to be aspects of the model or the way in
which some constructs from it are operationalized which require modification. To the extent that response time and HR were valid indicators of BAS activity in our studies, both reward and active avoidance cues clearly activate the BAS. Nonetheless, although both response time and HR showed substantial increases to active avoidance cues in experiment 2, surprisingly, they were essentially uncorrelated. It could be that these findings were idiosyncratic to our sample of criminal offenders. Because they are so counterintuitive, they will need to be replicated in a nonoffender sample before any firm conclusions can be drawn; however, it is also possible that one or both measures of BAS activation are contaminated by multiple influences when active avoidance as opposed to reward cues are present.

Regarding the BIS, response time and HR decreases to BIS cues appeared to be slightly more reliable indices of BIS activity than SCRs. Response time and HR showed decreases to BIS cues in both studies and both were correlated with self-reported anxiety in experiment 1. Participants did show significant increases in SC amplitude to punishment cues in both studies and SCRs increased to BIS cues in experiment 1; however none of the electrodermal indices of the BIS correlated with anxiety in either experiment. Our findings demonstrating a generally reliable increase in EDA to punishment cues are consistent with reports in the literature (Fowles, 1980, 1987). Additionally, the absence of a significant correlation between either electrodermal index of BIS activity and anxiety across both studies is consistent with Gomez and McLaren’s (1997) findings. The null relationship between our electrodermal results and anxiety contrasts with our finding in a recent study on criminal offenders where we found a greater number of SCRs following punishment in high-anxious compared with low-anxious offenders (Arnett, Howland, Smith & Newman, 1993).

It is possible that, given the gross motor demands of the task used in our current study, movement artifact within the electrodermal data may have introduced enough error to limit the reliability of these data and therefore attenuate any correlation they might show with other variables, such as our anxiety measure. In Arnett et al.’s (1993) study, where a positive relationship between anxiety and EDA was reported, motor demands for the tasks during which EDA was recorded were minimal. As noted earlier, Gomez and McLaren may have found null results between EDA and anxiety because their EDA measure — tonic SCL change — was less reliable than the measure used in the three studies cited above reporting positive findings. Arnett et al. (1993) used the average magnitude of SCRs; as noted earlier, this electrodermal measure has been shown to be a more reliable index of the BIS than tonic SCL (Fowles, 1980). Whatever the reason(s) for our weaker electrodermal results, the current study suggests that response time and HR decrease in the presence of punishment cues better index BIS influences than EDA increase.

Generally, the results from experiment 1 produced findings more consistent with the Gray/Fowles model than the results from experiment 2. As noted by Fowles (1980), it may be that because active avoidance cues are also cues for punishment, they generate some BIS activity and are therefore less ‘pure’ activators of BAS activation. Thus, in our study there was less of a contrast between our BAS and BIS phases in the active avoidance (experiment 2) compared with the reward study (experiment 1). Consistent with this interpretation, our post-hoc analysis showed a significant interaction between studies indicating that participants showed a
comparable number of SCRs during the mixed-incentive phases across both experiments, but more SCRs during the active avoidance phases (experiment 2) compared with the reward-only phases (experiment 1).

The finding that our indices of BAS activity, HR and response time, increased significantly in the presence of reward cues in experiment 1 contrasts with Gomez and McLaren’s (1997) failure to find significantly greater HR increase in their reward compared with their punishment condition. Because the amount of money participants won in the reward condition for both studies was comparable, it is unlikely that the BAS was activated more in our study than Gomez and McLaren’s (1997) study due to differences in the magnitude of reward used. We hypothesized earlier that one reason Gomez and McLaren (1997) failed to find significantly greater HR increase in their reward compared with their punishment condition may have been due to the intermittent presence of ‘frustrative non-reward’ in this condition that may have activated the BIS. Although frustrative non-reward was present in our reward condition as well, it was much less salient because participants were responding continuously and the rewards they received occurred, at most, once every five button presses. Thus, the decreased salience of frustrative non-reward in the ‘reward-only’ phase in our study most likely contributed to our positive findings relative to Gomez and McLaren’s (1997) null results. Another possible contributor to our positive findings relative to Gomez and McLaren’s (1997) null results may have involved the substantially greater motor demands in our task. It may be that the BAS is most activated when the organism is engaged in an approach set, especially one for reward, where motor systems are substantially activated. This speculation is consistent with three previous studies using a related continuous motor task which involved significant motor demands and 100% reward. In each case, significant increases in HR were associated with working for rewards (Fowles et al., 1982; Tranel, 1983; Tranel, Fisher & Fowles, 1982). Further supporting the association between reward cues and increased HR, these studies demonstrated a graded heart rate response according to the magnitude of the monetary incentive used; that is, greater incentives resulted in larger HR increases.

There are several limitations to our study. First, because we used only psychopathic and non-psychopathic criminal offenders as our sample, our results may not be generalizable to nonoffender populations. Therefore, replication of our findings with a nonoffender sample is clearly necessary. A second potential limitation concerns the possibility that the frequent motor movements involved in our task made it more difficult to detect spontaneous fluctuations related to underlying psychological processes such as anxiety. A task examining spontaneous fluctuations in SC in response to BIS cues may be more likely to reveal positive associations if the motor demands of the task are kept to a minimum. Nonetheless, our effects for SC amplitude to punishment cues were reasonably large and we still did not find that they were correlated with anxiety in either study. Finally, it may have been illuminating for us to employ anxiety and impulsivity measures that assess different components of these constructs to determine if different aspects of anxiety and impulsivity are related to the behavioral and physiological indices of the BIS and BAS we measured.

In conclusion, our investigation generally supported the validity of the Gray/Fowles model...
with some interesting exceptions. These exceptions are in need of further study and may have important implications for the model’s use in studying behavior, personality and psychopathology.

Acknowledgements

This research was supported by National Institute for Mental Health grant MH37711 to J.P.N. Presented in part at the 12th annual meeting of the Society for Research in Psychopathology, Palm Springs, CA, 1997. We wish to thank Randy Busse and Stevens Smith for their contributions to this research. Additionally, we thank the late Arnold Blahnik, Jeff Wydeven, Deb March, Bernice Conner and the staff of the Oakhill Correctional Institution and the Wisconsin Division of Corrections for their consistent cooperation and support regarding this research project.

References


