PERGAMON

Personality and Individual Differences 33 (2002) 511–532

PERSONALITY AND INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

J. A. Gray's reinforcement sensitivity theory: tests of the *joint subsystems hypothesis* of anxiety and impulsivity☆

Philip J. Corr*

Goldsmiths College, University of London, New Cross, London SE14 6NW, UK

Received 20 September 2000; received in revised form 6 August 2001

Abstract

Two experiments tested a new perspective on J. A Gray's Reinforcement Sensitivity Theory (RST) which postulates that the behavioural inhibition system (BIS) and the behavioural approach system (BAS) exert two separate effects on behaviour: (1) *facilitatory* (BIS–punishment, BAS–reward), and (2) *antagonistic* (BIS–reward, BAS–punishment). This *joint subsystems hypothesis* was contrasted with the conventional *separable subsystems hypothesis* of independent effects of the BIS and BAS in two paradigms: (1) affective modulation of the acoustic startle reflex ($n=70$), to measure the induction of emotional state; and (2) a visual information processing task with manipulations of reinforcement (feedback-alone vs. punishment of commission errors) and arousal (500 mg caffeine citrate vs. placebo; $n=120$), to measure behavioural inhibition/disinhibition. Consistent with the joint subsystems hypothesis: (1) high anxiety strengthened affective (electromyographic) reactions in the presence of unpleasant (compared with neutral) slides (i.e. fear potentiation), but this effect was stronger in low impulsivity participants (i.e. high impulsivity seemed to antagonise this BIS–mediated reaction); and (b) avoidance of punishment of incorrect responses was poorest in low anxiety, high impulsivity participants, pointing to a disinhibited pattern of reaction in individuals who, putatively, have a weak BIS and a strong BAS (this effect was found only in the caffeine group, suggesting that high levels of arousal may be necessary for the invigoration of disinhibitory behaviour). The implications of the joint subsystems hypothesis and the present data for Gray's reinforcement sensitivity theory of personality are discussed. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Punishment; Avoidance; Modulated startle reflex; Disinhibition; Arousal; Caffeine; Anxiety; Fear; Impulsivity; Behavioural inhibition; Behavioural approach

☆ A fuller version of this article is available from the author.

* Tel.: +44-20-7919-7885; fax: +44-20-7919-7873.
E-mail address: p.corr@gold.ac.uk

PII: S0191-8869(01)00170-2

1. General Introduction

J. A. Gray's (1970, 1987, 1991, 1994) Reinforcement Sensitivity Theory (RST) of personality is based upon reactions to rewarding and punishing stimuli in typical animal learning paradigms (e.g. conditioned emotional suppression of instrumental behaviour). This theoretical approach has spawned a number of alternative reinforcement-based theories of personality (e.g. Cloninger, 1986; Depue & Collins, 1999), and is now acknowledged as a major contribution to the neuropsychology of emotion and personality (Matthews & Gilliland, 1999). However, accumulating empirical evidence indicates that some of the basic predictions of RST are in need of clarification and perhaps reformulation (Corr, 2001). The aim of this article is to test one aspect of RST, *viz*. the independence of the reinforcement systems that underlie the personality traits of anxiety and impulsivity.

Gray's RST comprises three systems of emotion that underlie motivated behaviour.

The *fight/flight system* (FFS; Gray, 1987) was originally hypothesized to be sensitive to *unconditioned* aversive stimuli (i.e. innately painful stimuli), mediating the emotions of rage and panic (this system was aligned with the Eysencks' psychoticism factor, P; H. J. Eysenck & S. G. B. Eysenck, 1976). However, more recently, there has been a number of important revisions to this system (Gray & McNaughton, 2000). First, the concept of the FFS now incorporates freezing, which occurs in the presence of *actual* threat stimuli which are unavoidable (avoidable actual threat stimuli lead, depending on the situation, to either fear-related fleeing or anger-related fight). The FFS is now renamed the fight-flight-freezing system (FFFS; Gray & McNaughton, 2000, p. 86). Second, the FFFS mediates *all* aversive stimuli: unconditioned, innate and *conditioned*.

The *behavioural approach system* (BAS; Gray, 1987) was originally hypothesized to be sensitive to *conditioned* appetitive stimuli; the BAS forms a positive feedback loop, activated by the presentation of stimuli associated with reward and the termination/omission of signals of punishment. This system is responsible for positive affect. The BAS mediates impulsivity (Imp), which in terms of H. J. Eysenck's space ranges from E+/N+ (Imp+) to E−/N− (Imp−; Imp+ is rotated 30° from E; Gray, 1970; Pickering, Corr, & Gray, 1999).

In the revised theory, the BAS is sensitive to both conditioned and *unconditioned* appetitive stimuli. In this reformulation of the BAS, it is important to distinguish the incentive motivation component and the consummatory component of reactions to unconditioned appetitive stimuli. Gray still believes that no single system mediates the consummatory component of such reactions: for example, copulation and eating/drinking involve very different response systems. However, it may be assumed that the BAS is involved, to some extent, in moving the animal up the temporo-spatial gradient to the likely location of the primary reinforcer; that is, motivating the animal, by simple approach, to reduce the distance between current and desired appetitive state. The final act of consummatory behaviour would not be mediated by the BAS. As an example, consider typical human behaviour at lunch time: eliciting stimuli (e.g. clocks) signal the availability of appetitive stimuli (i.e. food/drink); these stimuli may trigger unconditioned physiological reactions which lead, with BAS activation, to locomotion to the location of food/drink (e.g. canteen) where the final unconditioned, consummatory (non BAS−mediated) act occurs.

The *behavioural inhibition system* (BIS; Gray, 1976, 1982) was originally hypothesized to be sensitive to *conditioned* aversive stimuli (i.e. signals of both punishment and the omission/termination of reward), extreme novelty, high intensity stimuli, and innate fear stimuli (e.g. snakes, blood). The BIS is the causal basis of anxiety (Anx), which in terms of H. J. Eysenck's personality space ranges from

E-/N+ (Anx+) to E+/N− (Anx−; Anx+ is rotated by 30° from N; Gray, 1970; Pickering et al., 1999). Upon activation, the BIS produces outputs of behavioural inhibition, an increase in arousal, heightened attention and information processing, and the emotion of anxiety.

Gray and McNaughton's (2000) revision of Gray's (1982) neuropsychology of anxiety has made important changes to the concept of the BIS. According to the revised theory, *conditioned* (along with unconditioned) aversive stimuli are mediated by the FFFS: the BIS is activated *only* when "...the animal's primary purpose is to achieve some goal which requires it to move *towards* a source of danger—that is, when it has *concurrent conflicting goals*: of reaching safety and of satisfying appetite" (Gray & McNaughton, 2000, p. 84; authors' own italics). Thus, "...the simplest way to activate the BIS is to concurrently activate the FFFS the BAS, i.e. face the animal with an approach avoidance conflict" (Gray & McNaughton, 2000, p. 86). Importantly, "...the presence of stimuli or contingencies *per se* is not sufficient to activate this system" (Gray & McNaughton, 2000, p. 86; authors' own italics). The approach-avoidance conflict elicits the state of anxiety; the presentation of an actual aversive stimulus, not involving approach, is mediated by the FFFS and corresponds to a separate state of fear. According to this revised theory, the BIS is now only activated when there is simultaneous activation of the BAS.

The implications for personality research of this new Gray and McNaughton (2000) theory has yet to be clarified. However, like Gray's (1982) original theory, this new version of RST would still seem to predict that, on average, impulsive (ex hypothesi, strong BAS) individuals should be most sensitive to signals of reward, relative to nonimpulsive (ex hypothesi, weak BAS) individuals; and anxious (ex hypothesi, strong BIS) individuals should be most sensitive to signals of punishment, relative to nonanxious (ex hypothesi, weak BIS) individuals. The proposed orthogonality of the BIS and BAS suggests that (1) responses to reward should be the same at all levels of anxiety, and (2) responses to punishment should be the same at all levels of impulsivity. This is the *separable subsystems hypothesis* of RST (Corr, 2001).

There is experimental support for RST's predicted association of introversion-extraversion and reinforcement (e.g. Boddy, Carver, & Rowley, 1986; Gupta, 1976, 1990; Gupta & Nagpal, 1978; Gupta & Shukla, 1989; Kantorowitz, 1978; McCord & Wakefield, 1981; Nagpal & Gupta, 1979; Seunath, 1975). But attempts to relate specific measures of Anx and Imp to Gray's reinforcement effects have met with only partial success (e.g. Gorenstein & Newman, 1980; Newman, 1987; for a review, see Matthews & Gilliland, 1999).

An extensive programme of research in Gray's own laboratory has found a plethora of mixed results (summarised in Pickering, Corr, Powell, Kumari, Thornton, & Gray, 1997). Some studies yielded positive support for punishment-mediated responses. For example, Corr, Pickering, and Gray (1997b) found superior procedural learning of anxious individuals under punishment, and inferior learning under a neutral condition; and electromyographic (EMG) startle reactions to unpleasant slides has been shown to be related to trait anxiety (Corr, Kumari, Wilson, Checkley, & Gray, 1997a; Corr et al., 1995b). However other studies have yielded inconsistent results. For example, Corr, Pickering, and Gray (1995a), using an instrumental learning task, with conditioned signals of reward and punishment, found that passive avoidance learning was related to Imp, not to Anx (i.e. Imp+ individuals showed impaired passive avoidance).

In the case of reward-mediated responses the situation is even more complex. For example, Anx− is sometimes found to moderate appetitive responses, whether assessed by EMG startle reactions to pleasant slides (e.g. Corr, et al., 1995b), induced positive emotion (Larsen & Katelaar, 1991), instrumental approach behaviour (Corr, Pickering, & Gray, 1995a), or appetitive classical conditioning (Mangan, 1978; Paisey & Mangan, 1988). Matthews and Gilliland (1999, Table 1, p. 600) showed that hedonic tone is consistently related to E+ and N− (i.e. Anx−), not Imp+ (i.e. E+/N+). Sometimes Imp is not found to be related to reward, either in terms of the learning of reward expectancies or behavioural responses to rewarding stimuli (e.g. Zinbarg & Mohlman, 1998).

In addition, complex Anx×Imp interactions are sometimes reported (e.g. Zinbarg & Mohlman, 1998; Zinbarg & Revelle, 1989). For example, Barratt (1971), using EEG, found that Imp+/Anx− individuals were less aroused at the moment of stimulus presentation, and emitted fewer classically conditioned responses, while Imp−/Anx+ individuals emitted the highest number of conditioned responses. Zinbarg and Mohlman (1998) concluded that "...the interactive effect of impulsivity by trait anxiety ... is not well understood at present" (p. 1038). Thus, there is limited support for the separable subsystems hypothesis, and some suggestion that the BIS and BAS jointly influence reward and punishment-mediated responses.

1.1. Joint subsystems hypothesis

In an attempt to account for the diversity of findings in the literature, Corr (2001) proposed a revision of RST to take into account the mutual interplay of BIS and BAS effects. Whereas the separable subsystems hypothesis states that "...individual differences in the functional capacity of one system are independent of the individual differences in the functional capacity of the other system" (Pickering, 1997, p. 145), the joint subsystems hypothesis postulates that the BIS and BAS have the *potential* to influence both reward-mediated and punishment-mediated behaviour; this qualifier relates to the operational parameters of reward and punishment which determine whether separable or joint effects are observed. Specifically, it is predicted that effects consistent with the separable subsystems hypothesis should be observed: (1) when strong appetitive/aversive stimuli are used; (2) when hyper-active BIS/BAS individuals are tested; and (3) in experimental situations that do not contain mixed reward and punishment cues, or demand rapid attentional and behavioural shifts between these two sets of motivational cues.

Given a background (non-zero) level of BIS/BAS activation, Anx (BIS) and Imp (BAS) may exert functionally interdependent effects. According to this joint subsystems hypothesis, (1) *state* appetitive responses and positive emotion should be highest in Imp+ (BAS+) and Anx− (BIS−) individuals; and (2) *state* measures of aversive responses and negative emotion should be highest in Anx+ (BIS+) and Imp− (BAS−) individuals. Where joint effects are in operation, behaviour should be observed as the statistical interaction of BIS/BAS effects (e.g. anxiety and impulsivity). The theoretical predictions of the separable subsystems hypothesis and the joint subsystems hypothesis are shown in Figs. 1 and 2, respectively.

The separable subsystems hypothesis and the joint subsystems hypothesis may be seen as complementary accounts within a two-process model of BIS/BAS functioning (Corr, 2001). It is argued that the BIS and BAS exert two effects, the first *facilitatory*, the second *antagonistic*. In the case of BIS behaviours, Anx+ facilitates, Imp+ antagonises; in the case of BAS behaviours, Imp+ facilitates, Anx+ antagonises. The precise pattern of personality effects is hypothesized to depend upon the relative strengths of aversive and appetitive stimuli (i.e. the degree of BIS and

BAS activation). With weak aversive and appetitive stimuli, only antagonistic factors may be at work, with Anx impairing BAS−mediated behaviour, and Imp impairing BIS−mediated behaviour; with strong stimuli, facilitatory factors may play a more important part, with Anx facilitating BIS−mediating behaviour, and Imp facilitating BAS−mediating behaviour. It is assumed that in the case of very strong BIS/BAS stimuli, antagonistic effects may be of little importance. A second factor of importance is assumed to be the value of Anx and Imp. If extreme personality groups are used, then results consistent with the separable subsystems hypothesis may be found because the subjective value of appetitive and aversive stimuli will be amplified in these individuals; in such groups, one system would effectively inhibit the alternate system and thus produce main effects of anxiety and impulsivity.

As it is much easier to manipulate aversive stimuli than appetitive stimuli in standard laboratory settings, it is perhaps not surprising that facilitatory effects of the BIS are more commonly found than facilitatory effects of the BAS; whereas, it is more common to find antagonistic effects of the BIS, antagonistic effects of the BAS are less common.

The causal mechanism underlying behavioural effects predicted by the joint subsystems hypothesis is assumed to be identical to the Gray–Smith (1969) arousal-decision model. It need only be assumed that

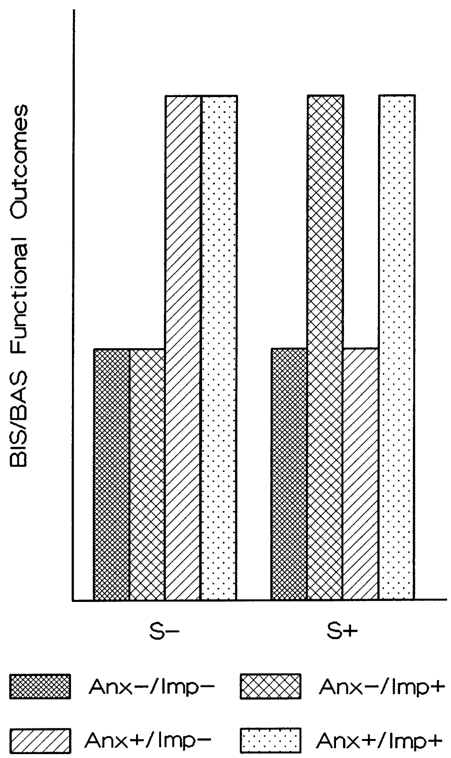

Fig. 1. *Separable subsystems hypothesis*: theoretical pattern of effects between low (−) and high (+) Anxiety (Anx) and Impulsivity (Imp) groups in strength of reaction to aversive (S−) and appetitive (S+) conditioned stimuli.

activation of the BIS/BAS (i.e. the punishment and reward mechanisms) does not totally inhibit activity of the alternate system; then it follows that the strength of BIS/BAS inputs to the decision mechanism would produce behavioural effects consistent with the joint subsystems hypothesis; *ex hypothesi*, resulting from the algebraic summation of BIS/BAS strengths entering the decision mechanism.[1]

Although the joint subsystems hypothesis was formulated on the basis of Gray's original (1976, 1982) BIS theory, Gray and McNaughton's (2000) revised BIS theory places BAS activation centre stage in BIS effects: *only* with BAS activation is the BIS activated (i.e. an approach-avoidance conflict; without BAS activation, aversive stimuli are mediated by the FFFS). Now,

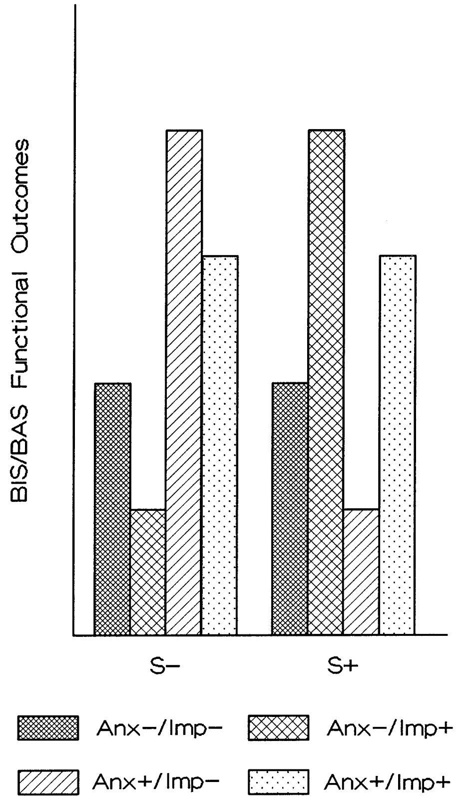

Fig. 2. *Joint subsystems hypothesis*: theoretical pattern of effects between low (−) and high (+) Anxiety (Anx) and Impulsivity (Imp) groups in strength of reaction to aversive (S−) and appetitive (S+) conditioned stimuli.

[1] Formally, assuming a linear model, BIS/BAS functional outcomes may be calculated: ((S−×BIS sensitivity)−(S+ x BAS sensitivity)), where S− is aversive stimulus strength, S+ appetitive stimulus strength. In terms of personality constructs, this may be reformulated: ((S−×Anx score)−(S+ x Imp score)). An assumption underlying this calculation of hypothesized BIS/BAS functional outcomes is that of a linear relationship in the inhibition of the BIS and BAS, such that a one unit increase/decrease in BIS/BAS activity leads to a comparable unit increase/decrease in the alternate system's influence. Empirically, this relationship may not be linear (e.g. the BIS may disproportionately inhibit the BAS). The precise pattern of BIS/BAS effects is an empirical question; importantly, the joint subsystems hypothesis does not rely upon this linear assumption: it is adopted for economy of exposition.

given this dependence of BIS activation upon BAS activation, the joint activation of these systems would seem central to understanding BIS effects, that is the joint subsystems hypothesis.

The joint subsystems hypothesis is also consistent with the claim that, "Conflict between the reward and punishment mechanism is resolved in the decision mechanism according to whichever input to this mechanism is stronger; the reciprocally inhibitory links between the reward and punishment mechanisms ensure a stable outcome to such conflicts" (Gray, 1987, p. 180). The algebraic summation of BIS/BAS strengths would ensure a dominant direction of response (albeit reduced in strength—i.e. antagonised—by input from the non-dominant system).

For the joint subsystems hypothesis to be viable, it need only be assumed that the output of the decision mechanism is the sum of inputs from *both* the BIS and the BAS: final behavioural effects being an algebraic summation of opposing reinforcement inputs. Therefore, the joint subsystems hypothesis is consistent Gray's theoretical model. It is assumed that, under typical conditions prevailing in the human psychological laboratory, BIS–BAS inhibitory effects are not as strong as those found in the typical animal laboratory where signals of reward and punishment are comparatively much stronger leading to greater BIS–BAS inhibition and behavioural outcomes more consistent with the separable subsystems hypothesis.

1.2. Aims and experimental predictions

Two separate experimental paradigms were employed to test the robustness and generalisability of the joint subsystems hypothesis: (1) affective modulation of the acoustic startle reflex; and (2) inhibited/disinhibited responding on an information processing task. These experimental measures were sufficiently distinct to ensure that similar results could not be due to a method artefact common to both paradigms.

2. Experiment 1: affective modulation of the acoustic startle reflex

2.1. Introduction

The startle reflex is a cross-species involuntary response to an abrupt, intense stimulus (e.g. a sudden noise), serving as a protective defence. In the rat, startle is measured by whole-body flinch; in human beings, it is easily and conveniently measured by the eyeblink reflex, measured electromyographically (EMG) from the orbicularis oculi (eyeblink) muscle.

A number of studies have shown that conditioned fear increases the magnitude of the startle reflex in animals (i.e. Davis, 1984). In human beings, affective modulation of the startle reflex is observed to increase in magnitude when the startle probes are presented during the viewing of unpleasant slides, as compared with neutral slides. Probes delivered during the viewing of pleasant slides significantly attenuate the startle, providing an independent measure of reactivity to pleasant events (i.e. Hamm, Greenwald, Bradley, & Lang, 1993). Furthermore, affective modulation is modality-free, consistent across different affective foreground stimuli (Bradley, Zack, & Lang, 1994), implying that it is mediated by central motivational states. Affective modulation of startle is seen to provide an objective measure of emotional reactivity in clinical (Cuthbert, Patrick, & Lang, 1991) and normal (Corr et al., 1995b) populations.

Affective startle modulation has also proven to be a useful tool to study the interrelation of normal traits of personality, emotion and emotion-related psychopathological conditions. Anxious individuals being more sensitive to threats may be expected to show greater potentiation of the startle reflex in the context of aversive stimuli, whereas impulsive individuals being more sensitive to incentives may display greater modulation of the reflex when exposed to pleasant stimuli. Corr et al. (1995b) found that Harm Avoidance, a measure of trait anxiety derived from Cloninger's (1986) neurobehavioural model of personality, significantly moderated fear potentiation (these effects were replicated by Corr, et al., 1997a). In addition, in Corr, et al. (1995b) there was evidence of an antagonism of anxiety on pleasure-attenuation.

2.1.1. Experimental predictions

The separable subsystems hypothesis predicts that: (1) affective reactions to unpleasant stimuli (i.e. fear potentiation) should be strongest in high anxiety participants, with no effect of impulsivity; and (2) affective reactions to pleasant stimuli (i.e. pleasure attenuation) should be strongest in high impulsivity participants, with no effect of anxiety.

In contrast, as the manipulation of the affective stimuli in this experiment was comparatively mild, and the experimental situation contained a mix of BIS/BAS stimuli, the joint subsystems hypothesis predicts effects of *both* anxiety and impulsivity: (1) affective reactions to unpleasant stimuli should be strongest in high anxiety, low impulsivity participants; and (2) affective reactions to pleasant stimuli should be strongest in high impulsivity, low anxiety, participants.

2.2. Method

2.2.1. Participants

Seventy volunteers, 30 males (mean age = 25.17, S.D. = 6.23) and 40 females (23.83, 6.89), were recruited from a university population; most participants took part for an exchange of course credits.

2.2.2. Design

A split-plot design was employed, with repeated measures taken on slide valence (pleasant, neutral and unpleasant slides); (median-split) anxiety and impulsivity groups served as the between-group factors.

2.2.3. Personality questionnaires

Impulsivity (Imp) was measured by the Impulsiveness (IVE) Questionnaire (which forms part of the Eysenck Personality Scales, EPS; H.J. Eysenck & S.G.B. Eysenck, 1991; mean = 8.50, S.D. = 4.54, range = 1–17). This scale measures non-planning (e.g. "Do you usually think carefully before doing anything?"; "Do you often get into a jam because you do things without thinking?") and behavioural impulsivity (e.g. "Do you often do things on the spur of the moment?"; "Do you often do things on impulse?"). This narrow measure of impulsivity is more appropriate for testing RST: Gray's impulsivity is correlated with extraversion and neuroticism, but also psychoticism. In addition, this scale has been widely used in the test of RST (Pickering et al., 1997).

The State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) provided the measure of trait anxiety (mean = 43.29, S.D. = 10.58, range = 20–66). The Pearson correlation between Anx and Imp was low and nonsignificant ($r = 0.20$ $P > 0.05$).

2.2.4. *Affective slide material*

Thirty-six slides were selected from the International Affective Picture System (IAPS; Lang, Öhman, & Vaitl, 1988). Twelve slides depicted unpleasant scenes (e.g. mutilated bodies), 12 neutral scenes (e.g. household objects), and 12 pleasant events (e.g. outdoor scenes). Each of the three slide valences were controlled for the arousal content of the slides (IAPS slide numbers: pleasant slides: 1750, 2540, 5830, 7580, 2530, 1460, 8080, 8030, 4660, 8200, 8490, and gender male/female specifics 4531/4003; neutral: 7010, 7050, 7000, 7150, 7090, 7031, 7820, 7830, 2230, 7620, 7190, 7560; unpleasant: 3230, 9180, 9000, 3160, 3300, 9265, 6230, 3031, 3000, 3170, 3150, 9410).

Each slide was presented for 8 s, followed by a randomly determined interval of 10–20 s (mean = 15). Startle probes were presented during 10 of the slides in each valence category; six startle probes were presented during the intervals in order to enhance unpredictability of the startle presentation (in total 36 probes were delivered).

Slides were arranged into six blocks, each of which consisted of two pleasant, two neutral, and two unpleasant slides. Each block contained one slide without a startle and one interval with a probe; the 6 blocks were ordered in a quasi-random sequence, as were the order of slides, intervals, and startle probes. For half of the participants this order of presentation was reversed.

2.2.5. *Equipment and scoring*

Stimulus presentation, data storing and recording, and procedural details were identical to those given in Corr, et al. (1995b). The acoustic startle probe comprised a 50-ms presentation of 100 dB (A) white noise with almost instantaneous rise time presented binaurally through headphones. A 3-min acclimatisation period [i.e. 70 dB (A) white noise only] was followed by six startle probes before the actual valence testing.

The dependent variable of interest was mean amplitude (of scorable responses, not all possible responses). Participants' data were used to compute amplitude scores if a minimum of four out of 10 possible reflexes were scorable.

2.2.6. *Procedure*

Participants were given written instructions, which included a brief rationale and procedure of the experiment, as well as details of physiological measurements and the presentation of unpleasant visual images. They were told that they could terminate the testing session at any time. Participants signed a consent form, and then completed the personality questionnaires. After a 3-min white-noise acclimatisation period, six preliminary startle probes were presented and then the slide presentation began. The experiment took place in a sound-attenuated room with dimmed lights.

2.3. *Results and discussion*

2.3.1. *Slide valence, order and gender effects*

A three-way (Valence×Order×Gender) ANOVA revealed a significant main effect of Valence, $F(2, 132) = 12.22$, $P < 0.001$ (linear component, $t = 3.79$, $P < 0.001$). As expected, a linear increase in amplitude was observed (Pleasant: $M = 142.33$, SEM = 16.49; Neutral: $M = 156.23$, SEM = 17.55; Unpleasant: $M = 171.47$, SEM = 18.54).

All other main and interaction effects were nonsignificant ($p > .05$), save an Order×Gender interaction, $F(1, 66) = 4.19$, $P < 0.05$ (this effect is of little theoretical significance, so is not considered further).

2.3.1.1. Effects of anxiety and impulsivity. An ANOVA, comprising Valence and (median-split) Anxiety (41.5) and Impulsivity (9), revealed a number of statistically and theoretically significant effects: (1) Valence×Anxiety, $F(2, 120) = 8.44$, $P < 0.001$ (linear component: $t = 3.22$, $P < 0.001$); (2) Valence x Impulsivity, $F(2, 120) = 5.66$, $P < 0.001$ (linear component: $t = 2.87$, $P < 0.01$); and (3) Valence×Anxiety×Impulsivity,$F(2, 120) = 2.33$, $P = 0.10$ (linear component: $t = 1.84$, $P = 0.07$; Fig. 3). The main effects of personality were not significant (Ps > 0.20) (The main effect of valence was not different to that already reported above.)

It is evident from Fig. 3 that, in the high anxiety group, there was a pattern of linear affective modulation of the startle reflex. A Valence×Impulsivity ANOVA, conducted separately for the low and high anxiety groups, confirmed this observation. In the low anxiety groups, the Valence×Impulsivity interaction was nonsignificant, $F(2, 62) = 0.42$, $P > 0.05$; in the high anxiety group it was significant, $F(2, 58) = 6.69$, $P < 0.01$. The crucial effect in this group was the stronger reaction

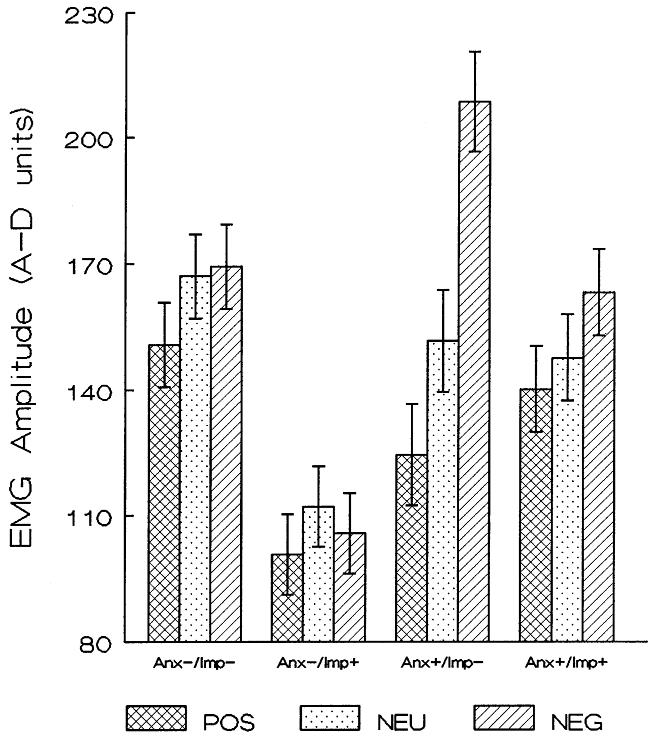

Fig. 3. Mean EMG eye blink amplitude scores for (median-split) low (–) and high (+) Anxiety (Anx) and Impulsivity (Imp) groups under positively (POS), neutrally (NEU) and negatively (NEG) valenced slide viewing conditions (bars = pooled within-groups standard error of mean differences).

in the low impulsivity group, as compared with the high impulsivity group: in the high anxiety, low impulsivity group the startle amplitude in the presence of unpleasant slides was statistically significant ($t=4.67$, $P<0.001$); in the low impulsivity group, this difference only approached a statistical trend ($t=1.56$, $P=0.15$); it is evident from the t values that the difference between these results is statistically significant. These effects support the prediction derived from the joint subsystems hypothesis, namely that affective reactions to unpleasant stimuli should be strongest in high anxiety, low impulsivity participants.

Thus, it appeared that high impulsivity, in anxious participants, acted as a form of protection from the induction of negative emotion (indeed the difference between amplitudes taken during the viewing of neutral and unpleasant slides only approached statistical significance in the high impulsivity group). This is the first study to show that the effects of high trait anxiety in the startle reflex is moderated by impulsivity.

Turning to modulation by pleasant slides, an unexpected pattern of effects was observed (Fig. 3). Statistically significant differences between pleasant and neutral slides were found for: (1) Anx−/Imp− ($t=2.15$, $P<0.05$); and (2) Anx+/Imp− ($t=2.60$, $P<0.05$); the Anx−/Imp+ ($t=1.66$, $P>.05$), and Anx+/Imp+ ($t=0.83$, $P>0.05$), were both nonsignificant.

Thus, affective modulation by pleasant slides was not related in a simple manner to (putatively BAS−mediated) impulsivity; in fact, from Fig. 3, it is evident that a statistically significant amplitude difference between neutral and pleasant slides was evident only for the low impulsivity groups, irrespective of level of anxiety; in the high impulsivity groups, there was no significant pleasure-attenuation of startle. Accordingly, this aspect of the data failed to support the second prediction, namely that affective reactions to pleasant stimuli should be strongest in high impulsivity, low anxiety, participants. However, as discussed elsewhere (Corr, et al., 1997a; Corr, et al., 1995b), it is likely that the acoustic startle reflex methodology is inherently biased against appetitive responses by virtue of the startle eliciting stimulus (i.e. the startle probe) being itself inherently aversive.

Comparing the actual results (Fig. 3) with the hypothesized results, derived from the joint subsystems hypothesis (Fig. 2), it is interesting to note that the amplitude differences between neutral and unpleasant slides replicates the pattern predicted by the joint subsystems hypothesis; namely, that fear potentiation would be strongest in the Anx+/Imp− group, next strongest in the Anx+/Imp+ group, followed by a weak effect in the Anx−/Imp−, and, lastly, an absence of fear potential in the group hypothesized to be least sensitive to aversive stimuli, *viz*., the Anx−/Imp+ (in this group the effects of the BIS seemed to have completely antagonised by, putatively, BAS−mediated Imp+). In contrast the pattern of observed effects was not consistent with predictions derived from the separable subsystems hypothesis, which predicted a main effect of anxiety in fear potentiation, and a main effect of impulsivity in pleasure-attenuation.

3. Experiment 2: inhibited/disinhibited reactions to punishing stimuli

3.1. Introduction

Reinforcement sensitivity theory was originally built upon the effects of conditioned stimuli for punishment in typical animal laboratory paradigms. Passive avoidance learning is a prototypical task for the measurement of reactions to punishment and anxiety (Gray, 1982). In a typical human laboratory experiment, *pure* passive avoidance tasks (i.e. tasks that are not influenced by other biological and cognitive processes) are not easy to identify; indeed, given the complexity of human behaviour, it may not be desirable to attempt to isolate passive avoidance from the range of other processes that are typically operative.

One popular experimental paradigm in cognitive and personality research is the rapid visual information processing (RVIP) task. This task has the virtue of containing two relatively separate processes, *viz*. (1) a target-sensitivity, detection (sensory) process (often measured by hit probability, or d' in signal detection analysis); and (2) a decision-criterion process (often measured by false alarms, β in signal detection analysis). Utilising this experimental paradigm, it is possible to separate these two processes, and thus gain a better understanding the dynamics underlying personality effects. In the present RVIP experiment, the effects of punishment on target sensitivity and decision-criterion were independently assessed.

The RVIP task was taken from Wesnes and Warburton (1984; a modified task originally used by Bakan, 1959), in which participants were required to detect 3 successive odd or even targets presented rapidly on a computer monitor. This task is engaging to the participant, requiring a non-zero level of operant responding, and it is sensitive to false alarms (i.e. errors of commission). In the present experiment, participants served under a feedback-alone or punishment of commission errors condition. In order to examine the putative effects of arousal (Gray & Smith, 1969), on target-sensitivity and decision-criterion parameters, participants were tested under either caffeine or placebo.

The operations of the BIS and BAS relate to information processing, as well as behavioural tendencies. Many apparently simple behavioural tasks entail a target/stimulus detection component, as well as a response-decision component. In these terms, the RVIP task may be conceptualised as a general analogue of processes that are common to many other cognitive and behavioural tasks; importantly, in the context of empirical tests of RST, the RVIP offers a convenient means by which to isolate target-sensitivity from decision-criterion components.

3.1.1. Experimental predictions

The separable subsystems hypothesis predicts that commission errors (i.e. the failure to inhibit a response to invalid target sequences) should be lowest in high anxiety participants under punishment; impulsivity should have no effect. Given that high anxiety individuals are typically high in arousal, caffeine might be expected to enhance these BIS effects.

In contrast, as the strength of the punishment manipulation was relatively mild (and, indeed, could be further weakened by not committing errors), the joint subsystems hypothesis predicts effects of *both* anxiety and impulsivity under punishment. There are two sets of possibilities. Commission errors will be either: (1) lowest in high anxiety, low impulsivity participants (i.e. Anx+/Imp−); or (2) highest in low anxiety, high impulsivity participants (i.e. Anx−/Imp+). This second possibility relates to the fact that this behavioural measure is bipolar; that is, it is sensitive: (1) to inhibition (i.e. withholding punishable responses); and (2) to disinhibition (i.e. making a high level of commission errors). Caffeine might be expected to enhance the effects of the BIS. Predictions derived from RST relating to the target-sensitivity parameter are less easy to specify; accordingly, these data were analysed and reported for completeness only (however, these data were important for showing the independence of target-sensitivity and decision-criterion processes in relation to the interactive effects of arousal, punishment and personality).

Level of arousal was manipulated because, according to Gray's model, although it should not alter the basic pattern of reinforcement effects, it may alter the intensity of behaviour. In the context of the joint subsystems hypothesis, this change in the intensity of behaviour may exert important effects on the pattern of interaction of anxiety and impulsivity, especially in passive avoidance tasks where there may be expected to be an antagonism between (1) the tendency to withhold responses (punishment-mediated) and (2) the tendency to emit more behaviour (arousal-mediated). From the experimental standpoint, the effects of arousal on punishment-mediated passive avoidance is unclear in human beings. One aim of this experiment was to explore this putative effect.

3.2. Method

3.2.1. Participants

One-hundred and twenty volunteers were recruited by placing advertisements in local newspapers, 60 males (mean age = 27.23, S.D. = 5.62), 60 females (28.23, 9.94). Participants received financial remuneration in exchange for their participation: in Feedback, they received £5.00; in punishment, they received £5.00 for taking part, plus what was left of an additional £5.00 gift from which they lost each time a commission avoidance (false alarm) was made.

3.2.2. Design

An independent-randomized groups design was used, comprising two levels of Arousal (caffeine citrate 500 mg, and placebo) and two levels of Reinforcement (feedback-alone, and punishment of commission errors). Participants were (quasi-) randomly allocated to one of the four Arousal × Reinforcement cells with the requirement that (1) equal numbers of males and females were in each condition, and (2) time of day did not become a systematic source of error. The only exclusion criterion was contraindications to caffeine (i.e. a personal or familial history of heart disease).

3.2.3. Personality questionnaires

Impulsivity (Imp) was measured by the Impulsiveness (IVE) Questionnaire (mean = 8.78, S.D. = 4.06, range = 1–18). The State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983) provided the measure of trait anxiety (Mean = 42.16, S.D. = 10.79, range = 20–71; see Section 2 for details of questionnaires). Anx and Imp were nonsignificantly correlated ($r = 0.15$, $P > 0.05$).

3.2.4. Caffeine administration

Two opaque gelatine capsules, containing either caffeine citrate 500 mg or inert white powder were emptied into a cup of blackcurrant drink (chosen to disguise the taste of caffeine citrate) in the presence of participants, who were blind to the Arousal condition.

3.2.5. Rapid visual information processing (RVIP) task

A modified version of Wesnes and Warburton's (1984) rapid visual information processing (RVIP) task was employed. It was of 25 min duration, divided into five blocks of 5 min. Each of the 5 blocks contained 500 digits (0–9), and 50 targets (i.e. for the whole experiment, 2500 digits and 250 targets were presented). The rate of stimulus presentation was one digit per 0.6 s (i.e. 100 digits per minute). The list of targets/nontargets was quasi-random and fixed for all participants. Targets were defined as three successive odd digits (e.g. 7, 1, 9) or three successive even digits (e.g.

6, 2, 8); "0" was designated as an even digit. From the onset of the last stimulus in the three-digit target sequence, 1.5 s was allowed for the participant to record a "hit"; only one hit was permitted in this time period.

Participants sat facing the computer monitor at a distance of approximately 60 cm; stimulus size was 2 cm at 1.5° visual angle. Stimuli were presented surrounded by a square box. Feedback and punishment messages were presented 3 cm above the stimulus box, and were presented for 2 s following a response (which was made by pressing a large red button on a purpose-built button box). A 2-min practice period preceded the main run, containing 200 digits and 20 targets (equal frequencies, and rate of stimulus presentation, as main run).

3.2.6. Instructions

Instructions for the practice period read:

> You will see on the computer monitor a long sequence of numbers. Your task is to detect sequences of three successive odd or even digits (for example, "2 8 6", "3 9 7", "0 8 2", 111 5 9"). As soon as you think you have seen one of these sequences press the button as fast as possible. Once you have responded, the computer will display "YES" to indicate that you have detected a sequence and "NO" to indicate that you have not detected a sequence. If you do not detect a sequence that had been presented then the computer will display "MISS" to inform you of this. *Attempt to be as accurate and fast as possible.*

Following the practice session, participants read one of the following two instructions. (Instructions were verbally reiterated to ensure that they were understood by all participants.)

> *Feedback*: For this part of the experiment, every time you detect a sequence the computer will display "CORRECT" to inform you of this. If you press the button when no sequence had been presented then the computer will display "INCORRECT". *Remember to be as accurate and fast as possible.*

> *Punishment of Commission Errors*. For this part of the experiment, you begin with £5.00. Every time you press the button when a sequence had not been presented you will lose 2 pence. The computer will display "INCORRECT" to inform you of this and show the amount of money you still have. If you detect a sequence the computer will display "CORRECT". *Remember to be as accurate and fast as possible.*

3.2.7. Equipment

The task was controlled by an ATARI ST 1040 microcomputer that registered and stored responses. The stimuli were presented centrally on an ATARI SC1224 monitor. A button box, consisting of a single button, was used for participants' responses.

3.2.8. Procedure

Participants were told that the experiment was concerned with task performance and personality, and that they would be required to take a drink that may contain caffeine (equivalent to three or four strong cups of coffee). Upon arrival, participants completed a consent form and caffeine/placebo was then administered. After 25 min (during which time the state and trait measures were taken), participants were introduced to the RVIP task, beginning with a 2-min practice before commencement of the main RVIP task (the rate of stimulus presentation and target frequency was identical to the main run). At the end of the task, the state measures were administered again, along with subjective reactions to the caffeine administration. (A full analysis of state effects may be obtained from the author.) Testing took place in a sound attenuated experimental cubicle.

3.2.9. Data scoring and analysis

For each 5-min period, the computer program computed: (1) the number of hits (i.e. correct detections); and (2) the number of false alarms (i.e. responses to non-target stimuli). Two dependent measures were thus available for analysis: (1) number of false alarms, and (2) hit probability. (The rationale for the decision to eschew a formal signal detection analysis in favour of this approach is given in the fuller version of this article, available from the author.)

Number of false alarms was used as the primary measure of behavioural inhibition/disinhibition (i.e. errors of commission); however, number of hits was also analysed for completeness. These data were examined for the presence of outliers, based on visual inspection of distributions in each arousal×punishment condition, looking for discontinuations in the data and using the method of identifying outliers described by Everitt (1996). Hit probability contained no obvious outliers, but false alarms did[2] (these data points were eliminated from analyses; it was evident during testing that these participants were randomly responding.) Next, the data were examined for normality: hit probability was normally distributed, but false alarms were highly positively skewed.[3] Following a logarithmic transformation (with 1 added to each data point to avoid zero values), false alarms were normally distributed. Logarithmic transformation of false alarms also dealt with the standard deviation being proportional to the mean. As is often recommended, inferential statistical analyses were based on transformed data; but descriptive presentation of these data is shown in original units.

First, two-way Arousal×Reinforcement ANOVAs were computed, separately for false alarms and hit probability, in order to test the effects of the two experimental factors; second, a four-way ANOVA was computed, comprising Arousal, Reinforcement, Anx and Imp. Personality scores were entered as continuous variables; personality scores×Arousal/Reinforcement factors were first transformed into interaction terms, and these new terms were entered into the analysis, providing a full linear model of all possible effects. This type of purpose-built analysis tests the homogeneity of slopes (not the means) of each personality term in the different experimental conditions, and is conceptually comparable to a regression analysis. Where permissable this method of inference testing is comparable to moderated multiple regression, and is preferable to taking median splits on the personality scales because of the preservation of statistical power (Cohen, 1968) and the reduction of statistical artefact (Bissonnette, Ickes, Berstein, & Knowles, 1990). Analyses were

[2] Cases eliminated included the following false alarm outliers (next included value in parentheses): Feedback, 443, 260, 204 (179); Punishment of commission errors, 321, 276, 122, 117 (78).

[3] Skewness values for false alarms (logged values in parentheses): Feedback, 2.587 (−0.434); Punishment of commission errors, 3.704 (−0.050).

performed in SPSS for Windows, using the MANOVA procedure. For ease of exposition, these data were graphed by taking medium splits on the personality scores, and showing the mean levels of performance in each experimental condition.

3.3. Results and discussion

Table 1 shows the number of false alarms, number of hits, and hit probability in the Arousal× Reinforcement conditions. Separate two-way ANOVAs on anxiety and impulsivity scores showed that these personality scores were not significantly different in Arousal or Reinforcement conditions (F-ratios for all main and interaction effects <1).

3.3.1. Arousal×Reinforcement×Blocks

Before examining the effects of personality on false alarms and hit probability taken across the whole experimental session, the temporal effects of Arousal and Reinforcement were examined by a Arousal×Reinforcement×Block ANOVA.

3.3.1.1. False alarms. Reinforcement, $F(1,109)=16.71$, $P<0.001$, was significant, relating to a much higher number of commission errors in feedback ($M=50.77$, S.E.M. $=6.67$) than punishment ($M=20.68$, S.E.M. $=2.89$). Thus, as expected, avoidance of incorrect responses was evident in punishment, reflecting the putative activation of the BIS. Neither the main effect of Arousal, $F(1, 109)=0.01$, $P>0.05$, nor the Arousal×Reinforcement interaction, $F(1,109)=0.82$, $P>0.05$, were significant.

The Block effect was also significant, $F(4, 436)=14.43$, $P<0.0001$ (Blocks: 1, M $=9.73$, S.E.M. $=0.97$; 2, 7.34, 0.81; 3, 6.01, 0.73; 4, 6.63, 0.92; 5, 6.15, 0.86). Post hoc tests revealed that there was a significant decline in false alarms during blocks 1–3, but no difference between blocks 3–5.

Effects involving Arousal×Reinforcement×Blocks were nonsignificant.

3.3.1.2. Hit probability. Reinforcement, $F(1, 109)=0.93$, $P>0.05$, Arousal, $F(1,109)=0.38$, $P>0.05$, and Arousal×Reinforcement interaction, $F(1,109)=1.60$, $P>0.05$, were nonsignificant. The only significant effect was Block, $F(4, 436)=8.02$, $P<0.001$ (Blocks: 1, $M=0.48$, S.E.M. $=0.02$; 2, 0.44, 0.02; 3, 0.46, 0.02; 4, 0.46, 0.02; 5, 0.48, 0.02). Post hoc tests revealed that the first and last block were identical, with hits declining sharply in block 2, recovering in blocks 3 and 4, and fully

Table 1
Means (S.D.) of the number of false alarms, number of hits, and hit probability under Arousal and Reinforcement conditions

	False alarms	Hits	Hits/targets
Feedback			
Caffeine	47.24 (51.55)	127.62 (30.93)	0.514 (0.124)
Placebo	54.43 (49.72)	116.61 (53.25)	0.451 (0.184)
Punishment of commission errors			
Caffeine	23.85 (24.84)	109.37 (49.90)	0.440 (0.201)
Placebo	17.72 (17.99)	114.49 (47.52)	0.461 (0.190)

recovering in block 5. This pattern of effects does not suggest a classic vigilance-decline in performance; accordingly it would have been problematic to include this Block factor in the analysis of personality effects because the underlying process generating this temporal fluctuation was not known.

3.3.2. *Personality effects*

Anxiety and impulsivity factors were added to the Arousal x Punishment ANOVA models.

3.3.2.1. False alarms. In addition to the main effect of Reinforcement already reported, the only significant effect was a four-way Arousal×Reinforcement×Imp×Anx, $F(1,94)=4.20$, $P<0.05$, term (all other terms did not approach significance, i.e. $Ps>0.10$). (This analysis was rerun with hit probability entered as a covariate, but this four-way interaction was unaffected.)[4]

To unravel this complex interaction, the sample was divided by Reinforcement group, and two separate three-way Arousal×Imp×Anx ANOVAs were run. In Feedback, the Arousal×Imp×Anx interaction was nonsignificant, $F(1,39)=0.01$, $P=0.96$; but in the punishment condition, the three-way interaction was significant, $F(1,39)=4.91$, $P<0.05$. Moreover, when the Imp×Anx interaction was decomposed further into placebo and caffeine conditions, no interaction

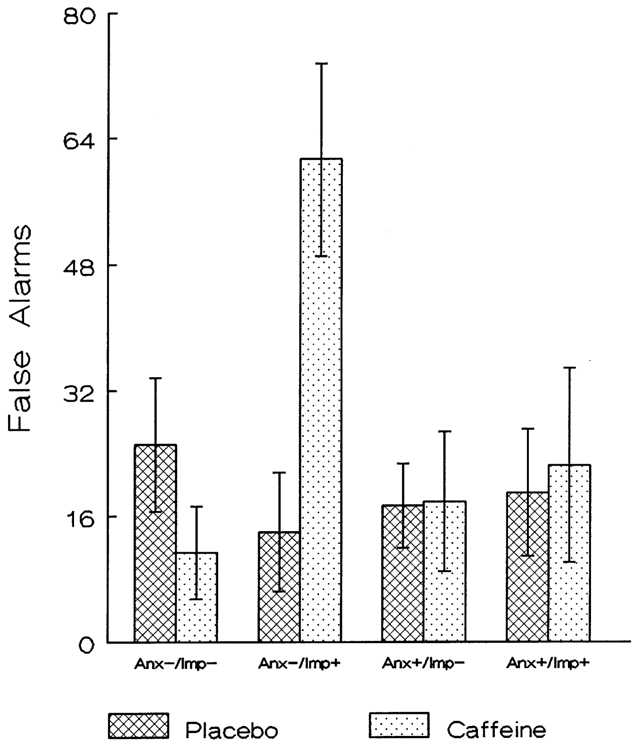

Fig. 4. Mean false alarms for (median-split) low (-) and high (+) Anxiety (Anx) and Impulsivity (Imp) groups in feedback-alone and punishment of commission errors, and caffeine and placebo conditions (bars = 1 standard error of the mean).

[4] An additional analysis was rerun to include the Block factor, but no additional significant effects involving this factor were observed, indicating that temporal dynamics were not operating in the Punishment×Arousal×Anxiety×Impulsivity interaction.

was evident under placebo, $F(1, 20) = 0.63$, $P = 0.44$, but the interaction under caffeine was significant, $F(1, 17) = 4.57$, $P < 0.05$. Fig. 4 shows that the (median-split) Anx−/Imp+ group under caffeine had the largest number of commission errors; these data compare with a higher mean level of commission errors in Feedback that hovered around 50 (Table 1). (Reanalysis with all data points included revealed the same pattern of effects; the Anx−/Imp+ group continued to show the largest number of false alarms under caffeine.)

3.3.2.2. Hit probability. In addition to the Reinforcement main effect already reported, no other significant, or near-significant (i.e. all Ps > 0.10), effects were found (footnote 4).

Fig. 4 shows that low anxiety, high impulsivity (Anx−/Imp+) participants in the caffeine condition had very poor avoidance of punishable stimuli, incurring a greater number of false alarms than any of the other personality groups. This pattern of disinhibited behaviour is consistent with the view that anxiety imposes a brake on impulsive behaviour (i.e. anxiety facilitates BIS−medimediated responses), and impulsivity antagonises BIS−mediated responses.

Apart from the Anx−/Imp+ group in caffeine, all other groups showed effective behavioural inhibition, and therefore levels of punishment-induced arousal were low. It thus seems that caffeine-induced arousal in impulsive individuals (i.e. Imp+), who also had no effective brake on inappropriate responses (i.e. Anx−), led to greater behavioural activation which in turn impaired behavioural inhibition. In most real-life punishment situations, it may be assumed that punishment-induced arousal would be considerably greater than that seen in the present experiment. Therefore, it could be concluded that the induction of arousal increases the probability of the emission of impulsive responding in vulnerable individuals (i.e. Anx−/Imp+). In a mixed punishment-reward context, it may further be assumed that appetitive motivation would lead to more commission errors in high impulsivity participants (especially, if they are also low in anxiety) by virtue of arousal-induction. This Newman-type pattern of results (Newman, 1987) is sometimes found (e.g. Corr, Pickering, & Gray, 1995a). In the present experiment, perhaps the induction of arousal by punishment alone was too weak to set in train the processes leading to disinhibitory responses.

Given the absence of punishment, arousal and personality effects on hit probability, the possibility that false alarms were a secondary effect to a primary effect on target-sensitivity may be ruled-out. Therefore, it may be concluded that the personality effects observed related to behavioural tendencies (decision-criterion) rather than target-detection (sensory sensitivity).

Comparing the actual results (Fig. 4) with the hypothesized results, derived from the joint subsystems hypothesis (Fig. 2), it is interesting to see that the most disinhibited participants belonged to the Anx−/Imp+ group which the joint subsystems hypothesis predicted would be least sensitive to aversive stimuli and most sensitive to appetitive stimuli. It is likely that there were implicit appetitive stimuli in the experimental setting; this may have been sufficient to activate their BAS, adding to their existing tendency towards disinhibition by virtue of their relative under-active BISs.

These data suggest that, with relatively weak punishment, disinhibitory effects (*ex hypothesi*, antagonistic effects of the BAS on behavioural inhibition) are observed (1) when there is no effective brake on impulsive behaviour (i.e. weak BIS, Anx−), and (2) with an sufficient level of arousal, that increases the probability of the emission of behaviour (Duffy, 1962).

Newman (1987) argued that impulsive individuals show impaired passive avoidance because of a response modulation deficit resulting from their greater sensitivity to reward. The present data indicate that such disinhibition may occur in the absence of overt reward, and indeed may be the result of response invigoration (in the present experiment, induced by caffeine; in other experiments, external manipulations of reward). It is thus possible that disinhibited behaviour may be *initiated* by reward responsivity by virtue of the induction of arousal. However, it could be argued that, in the punishment condition, there were implicit signals of reward (i.e. correct hits resulting in positive feedback), therefore, it is not possible to rule out entirely the possibility that the disinhibited behaviour observed was a direct consequence of a response modulation deficit of the type postulated by Newman (1987). Further experiments are needed to address this question.

4. General discussion

The purpose of this article was to provide an empirical contrast of two positions with respect to Gray's Reinforcement Sensitivity Theory (RST), *viz*. the conventional separable subsystems hypothesis and the newer joint subsystems hypothesis (Corr, 2001). The two experiments yielded results that were consistent with the joint subsystems hypothesis; predictions based on the separable subsystems hypothesis were not confirmed. In the two experiments, interactive effects of Anxiety×Impulsivity in reactions to manipulations of punishment were observed; and the pattern of effects from these separate studies were complementary.

In Experiment 1, Imp+ reduced (i.e. antagonised) Anx+ participants' EMG reactions in the presence of unpleasant slide material. With respect to this fear potentiation, the pattern of effects observed (Fig. 3) closely resembled the predictions based on the joint subsystems hypothesis (Fig. 2). Indeed, the magnitude of the amplitude difference between neutral and unpleasant slides exactly matched the hypothesized state effects of joint BIS/BAS strengths: Anx+/Imp− participants were most reactive to unpleasant slides, Anx−/Imp+ least reactive; and, significantly, for the joint subsystems hypothesis, the Anx+/Imp+ participants showed attenuated (i.e. antagonised) fear potentiation. In Experiment 2, false alarms were highest in Imp+/Anx− participants, suggesting that high anxiety antagonised the disinhibitory effects of impulsivity.

It is possible that the experiments which have revealed joint effects of anxiety and impulsivity might have resulted from the tasks containing mixed implicit signals of reward and punishment, thus concealing the truly separable and independent effects of the BIS and BAS. In Experiment 1, both pleasant and unpleasant slide material were presented; and in Experiment 2, the word "correct" was presented after each correct response. Thus, in both experiments, there was a mixture of aversive and appetitive stimuli. The separable subsystems hypothesis may be confirmed on tasks that do not contain mixed reinforcement.

In typical human personality research, it may be unrealistic to assume that one reinforcement system dominates over the other, therefore, in laboratory studies, the joint subsystems hypothesis may be most appropriate for understanding individual differences in reactions to reinforcing stimuli. In terms of everyday life, this problem may be compounded as most personal and social situations entail a mixture of appetitive and aversive stimuli. Even in situations which can be objectively defined as aversive, cognitive appraisal and the *potential* of a favourable outcome may be sufficient to activate the BAS, and this may lead to effects consistent with the joint subsystems hypothesis. Accordingly, one major distinction between the separable subsystems hypothesis and the joint subsystems hypothesis is their relevance to *ideal* and the *actual* experimental situations.

Closer attention to actual reinforcing stimuli seems sensible in order to refine operational definitions, thus potentially enhancing the experimental precision of RST (Corr, 2001).

In future tests of the separable and joint subsystems hypotheses, effects of level of arousal should be examined. Arousal effects might be especially important on tasks where there are opposing motivational tendencies of (1) withholding (punishment-mediated) responses, and (2) greater (arousal-mediated) behavioural intensity.

Discussion of the functional outcomes determined by the decision mechanism raises an intriguing issue concerning the structural nature of personality. On the one hand, there is considerable statistical support for the Eysencks' structural model of personality (especially, extraversion and neuroticism; factors which are found in virtually all other structural models; e.g. the five-factor model); on the other hand, there is now accumulating evidence for Gray's RST processes. Is it possible to unite the Eysenck's structural model with Gray's causal model? One possibility is that the joint effects of the BIS and BAS give rise to statistical factors of extraversion and neuroticism, *even though* anxiety and impulsivity may be more closely aligned to the BIS and BAS, respectively. By subtracting the values shown in Fig. 2, thus yielding combined BIS/BAS scores, we find the following results: strongest BAS influence in Imp+/Anx− (i.e. extraversion), strongest BIS influence in Anx+/Imp− (i.e. introversion). This is essentially the position adopted by Newman's model of personality (i.e. introversion–extraversion is the *balance* of BIS–BAS sensitivities). Neuroticism is then left to reflect the *strength* of BIS–BAS activation. (Assuming that strong activation of the BIS exerts disproportionate suppression on the influence of the BAS—a key concept in RST—then neuroticism would be more associated with negative affect, extraversion being associated with positive affect). This suggestion, of course, may be traced to Gray's (1970) original formulation of RST; what is different in terms of the joint subsystem hypothesis is the suggestion that the statistical investigations of the *surface expression* of personality (i.e. statistical model) may well be sensitive to the interactive effects of underlying BIS/BAS systems (i.e. causal model). The Gray and McNaughton (2000) revision of Gray's (1982) original theory now also stresses the joint involvement of the BIS and the BAS in anxiety. Accordingly, there may be no real contradiction between the Eysencks' extraversion and neuroticism (descriptive) factors, and Gray's BIS/BAS (causal) factors.

In conclusion, in two separate experiments, interactions of anxiety and impulsivity were observed on punishment-related processes, supporting the contention of the joint subsystems hypothesis of BIS/BAS effects. This hypothesis contrasts with conventional tests of Gray's RST, which assumes that the BIS and BAS exert separate and independent effects, that is, the separable subsystems hypothesis. These results cast further doubt upon the validity of the separable subsystems hypothesis in *typical* human experimental contexts (however, it may be valid in experimental contexts that contain either (1) strong appetitive/aversive stimuli, or (2) extreme personality groups). Although the veridicality of the joint subsystems hypothesis must await further experimental investigation, it suggests a possible explanation for the variety of anxiety and impulsivity effects observed in RST studies.

Acknowledgements

Two anonymous referees provided valuable comments on an earlier draft of this article. I wish to thank Jeffrey A. Gray and Alan Pickering for many insightful discussions concerning reinforcement sensitivity theory (RST); and to Jeffrey A. Gray and Neil McNaughton for clarification of the Gray-McNaughton (2000) revision of RST. I acknowledge Atsuko Inoue for assistance with the collection of the data reported in Experiment 1.

References

Bakan, P. (1959). Extroversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology*, *50*, 325–332.

Barratt, E. S. (1971). Psychophysiological correlates of classical differential eyelid conditioning among normal subjects selected on the basis of impulsiveness and anxiety. *Biological Psychiatry*, *3*, 339–346.

Bissonnette, V., Ickes, W., Berstein, I., & Knowles, E. (1990). Personality moderating variables: a warning about statistical artefact and a comparison of analytic techniques. *Journal of Personality*, *58*, 657–687.

Boddy, J., Carver, A., & Rowley, K. (1986). Effect of positive and negative reinforcement on performance as a function of introversion-extraversion: some tests of Gray's theory. *Personality and Individual Differences*, *7*, 81–88.

Bradley, M. M., Zack, J., & Lang, P. J. (1994). Cries, screams, and shouts of joy: affective responses to environmental sounds (Abstract). *Psychophysiology, 31, S29*..

Cloninger, C. R. (1986). A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatric Developments*, *3*, 167–226.

Cohen, J. (1968). Multiple Regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426–443.

Corr, P. J. (2001). Testing problems in J. A. Gray's personality theory: a commentary on Matthews and Gilliland (1999). *Personality and Individual Differences*, *30*, 333–352.

Corr, P. J., Kumari, V., Wilson, G. D., Checkley, S., & Gray, J. A. (1997a). Harm avoidance and affective modulation of the startle reflex: a replication. *Personality and Individual Differences*, *22*, 591–593.

Corr, P. J., Pickering, A. D., & Gray, J. A. (1995a). Personality and reinforcement in associative and instrumental learning. *Personality and Individual Differences*, *19*, 47–71.

Corr, P. J., Pickering, A. D., & Gray, J. A. (1997b). Personality, punishment, and procedural learning: a test of J. A. Gray anxiety theory. *Journal of Personality and Social Psychology*, *73*, 337–344.

Corr, P. J., Wilson, G. D., Fotiadou, M., Kumari, V., Gray, N. S., Checkley, S., & Gray, J. A. (1995b). Personality and affective modulation of the startle reflex. *Personality and Individual Differences*, *19*, 543–553.

Cuthbert, B. N., Patrick, C. J., & Lang, P. J. (1991). Imagery in anxiety disorder patients: visceral and startle probe responses. *Psychophysiology*, *28*, S18 (Abstract)..

Davis, M. (1984). The mammalian startle response. In R. C. Eaton (Ed.), *Neural mechanisms of startle behaviour* (pp. 287–351). New York: Plenum Press.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, *22*, 491–533.

Duffy, E. (1962). *Activation and behaviour*. London: Wiley.

Everitt, B. S. (1996). *Making sense of statistics in psychology*. Oxford: Oxford University Press.

Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. London: Hodder and Stoughton.

Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales*. London: Hodder and Stoughton.

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: a new perspective and a model for research. *Psychological Review*, *87*, 301–315.

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy*, *8*, 249–266.

Gray, J. A. (1976). The behavioural inhibition system: a possible substrate for anxiety. In M. P. Feldman, & A. M. Broadhurst (Eds.), *Theoretical and experimental bases of behaviour modification* (pp. 3–41). London: Wiley.

Gray, J. A. (1982). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system (first edition)*. Oxford: Oxford University Press.

Gray, J. A. (1987). *The psychology of fear and stress*. Cambridge: Cambridge University Press.

Gray, J. A. (1991). Neural systems of motivation, emotion and affect. In J. Madden (Ed.), *Neurobiology of learning, emotion and affect* (pp. 273–306). New York: Raven Press.

Gray, J. A. (1994). Framework for a taxonomy of psychiatric disorder. In S. H. M. van Goozen, N. E. van de Poll, & J. A. Sergeant (Eds.), *Emotions: essays on emotion theory* (pp. 29–59). Hove: Lawrence Erlbaum.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system (second edition)*. Oxford: Oxford University Press.

Gray, J. A., & Smith, P. T. (1969). An arousal decision model for partial reinforcement and discrimination learning. In R. M. Gilbert, & N. S. Sutherland (Eds.), *Animal discrimination learning* (pp. 243–272). London: Academic Press.

Gupta, B. S. (1976). Extraversion and reinforcement in verbal operant conditioning. *British Journal of Psychology*, *67*, 47–52.

Gupta, B. S. (1990). Impulsivity/sociability and reinforcement in verbal operant conditioning: a replication. *Personality and Individual Differences*, *11*, 585–589.

Gupta, B. S., & Nagpal, M. (1978). Impulsivity/sociability and reinforcement in verbal operant conditioning. *British Journal of Psychology*, *68*, 203–206.

Gupta, B. S., & Shukla, A. P. (1989). Verbal operant conditioning as a function of extraversion and reinforcement. *British Journal of Psychology*, *80*, 39–44.

Hamm, A. O., Greenwald, M. K., Bradley, M. M., & Lang, P. J. (1993). Emotional learning, hedonic change, and the startle probe. *Journal of Abnormal Psychology*, *102*, 453–465.

Kantorowitz, D. A. (1978). Personality and conditioning of tumescence and detumescence. *Behaviour Research and Therapy*, *16*, 117–123.

Lang, P. J., Öhman, A., & Vaitl, D. (1988). *The international affective picture system (photogragphic slides)*. Gainsville, FL: University of Florida (Centre for Research in Psychophysiology).

Larsen, R. J., & Katelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology*, *61*, 132–140.

Mangan, G. L. (1978). Factors of conditionability and their relationship to personality types. *Pavlovian Journal of Biological Science*, *13*, 226–235.

Matthews, G., & Gilliland, K. (1999). The personality theories of H. J. Eysenck and J. A Gray: a comparative review. *Personality and Individual Differences*, *26*, 583–626.

McCord, R. R., & Wakefield, J. A. (1981). Arithmetic achievement as a function of introversion-extraversion and teacher-presented reward and punishment. *Personality and Individual Differences*, *2*, 145–152.

Nagpal, M., & Gupta, B. S. (1979). Personality, reinforcement and verbal operant conditioning. *British Journal of Psychology*, *70*, 471–476.

Newman, J. P. (1987). Reaction to punishment in extraverts and psychopaths: implications for the impulsive behaviour of disinhibited individuals. *Journal of Research in Personality*, *21*, 464–480.

Paisey, T. J. H., & Mangan, G. L. (1988). Personality and conditioning with appetitive and aversive stimuli. *Personality and Individual Differences*, *9*, 69–78.

Pickering, A. D. (1997). The conceptual nervous system and personality: from Pavlov to neural networks. *European Psychologist*, *2*, 139–163.

Pickering, A. D., Corr, P. J., & Gray, J. A. (1999). Interactions and reinforcement sensitivity theory: a theoretical analysis of Rusting and Larsen (1997). *Personality and Individual Differences*, *26*, 357–365.

Pickering, A. D., Corr, P. J., Powell, J. H., Kumari, V., Thornton, J. C., & Gray, J. A. (1997). Individual differences in reactions to reinforcing stimuli are neither black nor white: to what extent are they Gray?. In H. Nyborg (Ed.), *The scientific study of human nature: tribute to Hans J. Eysenck at eighty* (pp. 36–67). London: Elsevier Sciences.

Seunath, O. M. (1975). Personality, reinforcement and learning. *Perceptual and Motor Skills*, *41*, 459–463.

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

Wesnes, K., & Warburton, D. M. (1984). Effects of scopolamine and nicotine on human rapid visual information processing. *Psychopharmacology*, *82*, 147–150.

Zinbarg, R. R., & Mohlman, J. (1998). Individual differences in acquisition of affectively valenced associations. *Journal of Personality and Social Psychology*, *74*, 1024–1040.

Zinbarg, R., & Revelle, R. (1989). Personality and conditioning: a test of four models. *Journal of Personality and Social Psychology*, *57*, 301–314.