Pickering, Alan and Corr, Philip

J.A.Gray's reinforcement sensitivity theory (RST) of personality

11

J.A. Gray's Reinforcement Sensitivity Theory (RST) of Personality

Alan Pickering and Philip Corr

Jeffrey Gray's (1976, 1982) behavioural inhibition system (BIS) theory of anxiety has stood well the test of time. This theory of personality – which is now widely known as *reinforcement sensitivity theory* (RST) – has gradually evolved over the past 30 years, seeing its major revision in 2000 by Gray and McNaughton, and even further elaborations and refinements subsequently (McNaughton and Corr, 2004, 2008; Corr and McNaughton, 2008). However, recent data that have strengthened the general foundations of the neural basis of the theory have also forced significant modifications of, and additions to, its superstructure. These changes are not inconsequential; as such, predictions cannot now be based on prior knowledge of the 1982 version. These changes, we contend, have the potential to lead to confusion. A major purpose of this chapter is to review the current scientific status of Gray's RST and draw out some of its major implications for future research.

RST is built upon a *state* description of neural systems and associated, relatively short-term, emotions and behaviours, which, according to the theory, give rise to longer-term *trait* dispositions of emotion and behaviour. This theory argues that statistically defined personality factors are sources of variation that are stable over time and that derive from underlying properties of an individual; it is these, and current changes in the environment, that comprise the neuropsychological foundations of 'personality'. This assertion is demanded by the fact that personality traits account for behavioural differences between individuals presented with identical environments; also, behavioural differences show consistency across time. Thus, the ultimate goal of personality research is to identify the relatively static (underlying) biological variables that determine the (superficial) factor structure measured in behaviour. It would, of course, be a mistake to deny the relevance of the environment in controlling behaviour, but to produce consistent long-term effects, environmental influences must be mediated by, and instantiated in, biological systems.

Gray's approach to the biological basis of personality followed a particular pattern: (a) first identify the fundamental properties of brain-behavioural systems that might be involved in the important sources of variation observed in human behaviour and (b) then relate variations in these systems to known measures of personality. Central to this approach is the assumption that the variation observed in the functioning of these brain-behavioural systems comprise what we term 'personality'. As discussed below, relating (a) to (b) has proved the major challenge to RST researchers.

Now, most RST studies have tested the unrevised (pre-2000) version of RST. But, as we shall see, in many crucial respects, the revised Gray and NcNaughton (2000) theory of the underlying neural systems and their function is very different, leading to the formulation of new personality hypotheses, some of which stand in opposition to those generated from the unrevised theory (for more detailed discussion of these matters, see Corr, 2004, 2008; Corr and McNaughton, 2008; McNaughton and Corr, 2004, 2008).

'CLASSIC' (1970–2000) AND REVISED (2000–) REINFORCEMENT SENSITIVITY THEORY

Today, in personality research, it is common to relate personality factors to emotion and motivational systems, but this consensus did not prevail before the time of Gray's original work. It is a mark of achievement that Gray's (1970, 1982) approach is today so widely accepted, and the emergence of a *neuroscience of personality* can be seen to be largely shaped by his work. In a similar vein to Hans Eysenck's (1957, 1967) theories before him, Gray's innovation was to put together the existing pieces of the scientific jigsaw in order to provide the foundations of a general theory of personality. Gray, like Pavlov (1927) before him, advocated a twin-track approach: the *conceptual nervous system* (cns), and the *central nervous system* (CNS) (cf. Hebb, 1955). That is, the cns components of personality (e.g. learning theory; see Gray, 1975) and the component brain systems underlying systematic variations in behaviour (ex hypothesi, personality). As noted by Gray (1972a), these two levels of explanation *must* be compatible, but given a state of imperfect knowledge it would be unwise to abandon one approach in favour of the other. Gray used the language of cybernetics, in the form of cns–CNS bridge, to show how the flow of information and control of outputs is achieved (e.g. the Gray and Smith, 1969, 'arousal-decision' model).

Theoretical origins of RST

In contrast to Gray's bottom-up general approach, Hans Eysenck adopted a very different 'top-down' method. His search for causal systems was determined by the structure of statistically derived personality factors/dimensions. In an important respect, Eysenck's approach was viable: this was to understand the causal bases of *observed* personality structure, defined as a unitary whole (e.g. extraversion and neuroticism). For this very reason, it is perhaps not surprising to learn that Eysenck's causal systems never developed beyond the postulation of a small number of very general brain processes, principally the ascending reticular activating system (ARAS), underlying the dimension of introversion–extraversion and cortical arousal (for a summary see Corr, 2004). A second dimension, neuroticism (N), was related to activation of the limbic system and emotional instability (see Eysenck and Eysenck, 1985). Taken together, Gray's and Eysenck's approaches are complementary, tackling important problems at different levels of analysis.

Eysenck's (1967) arousal theory of extraversion hypothesized that introverts and extraverts differ with respect to the sensitivity of their cortical arousal system; and this is in

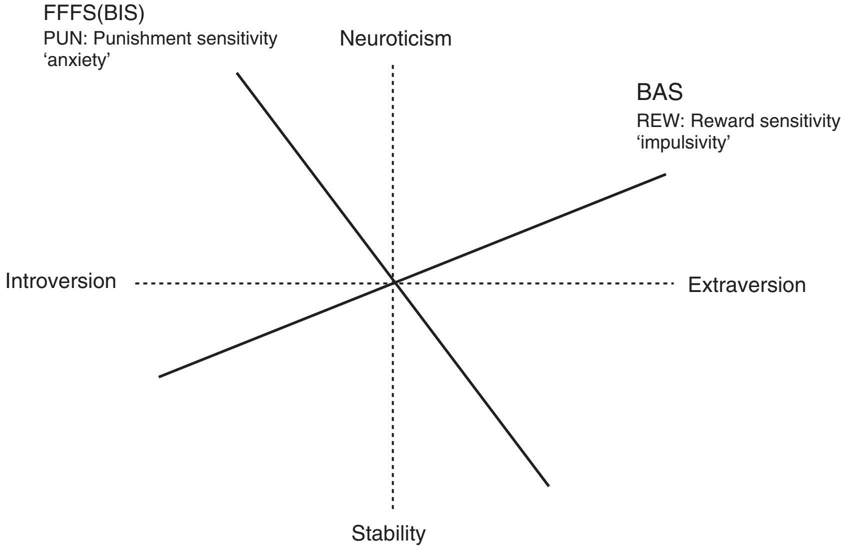

Figure 11.1 Position in factor space of the fundamental punishment sensitivity and reward sensitivity (unbroken lines) and the emergent *surface expressions* of these sensitivities, viz. extraversion (E) and neuroticism (N) (broken lines). The current working hypothesis is that 'punishment sensitivity' – which, in the unrevised model, was labelled 'anxiety – relates to both the FFFS and BIS'

consequence of differences in response thresholds of their ARAS. According to this theory, compared with extraverts, introverts have lower response thresholds and thus higher cortical arousal. In general, introverts were said to be more cortically aroused and more arousable when faced with sensory stimulation. However, the extraversion-arousal champions marched under a banner upon which was blazoned an inverted-U symbol – chosen, in large measure, by virtue of the Pavlovian notion of transmarginal inhibition (TMI; a protective mechanism that breaks the link between increasing stimuli intensity and behaviour at high intensity levels – in the Hullian learning literature this effect went under the name of 'stimulus intensity dynamism'). It was against this theoretical backdrop that RST developed.

Gray's (1970, 1972b, 1981) modification of Eysenck's theory proposed changes: (a) to the position of extraversion (E) and neuroticism (N) in Eysenckian factor space; and (b) to their neuropsychological bases. Gray argued that E and N should be rotated by approximately 30 degrees to form the more causally efficient axes of 'punishment sensitivity', reflecting anxiety (Anx), and 'reward sensitivity', reflecting impulsivity (Imp) (Figure 11.1; see Pickering et al., 1999).

This modification stated that Imp+ individuals are more sensitive to *signals* of reward, relative to Imp– individuals, and Anx+ individuals are more sensitive to *signals* of punishment, relative to Anx– individuals. The proposed independence of the axes suggested that (a) responses to reward should be the same at all levels of Anx and (b) responses to punishment should be the same at all levels of Imp – this position was dubbed the 'separable subsystems hypothesis' by Corr (2001, 2002). According to RST, Eysenck's E and N dimensions are derivative secondary factors of these more fundamental punishment and reward sensitivities: E reflects the balance of punishment and reward sensitivities; N reflects their joint strengths (Gray, 1981).

Clinical neurosis

Eysenck's taxonomic model of personality was based on the factor analysis of the symptoms of war 'neurotics' (1944, 1947), and his 1957 and 1967 causal theories were designed to explain the genesis of these neuroses; it is, thus, on these grounds that the theory is critically tested. In brief, Eysenck postulated that introverts are more prone to suffer from anxiety disorders by virtue of their greater conditionability, especially of emotional responses. This theory was later elaborated to include the notion of incubation effects in conditioning (Eysenck, 1979), in order to account for the 'neurotic paradox' (i.e. the failure of extinction with continued non-reinforcement of the CS). Coupled with emotional instability, reflected in N, this made the introverted neurotic (E−/N+) particularly prone to anxiety disorders.

However, from the very beginning of this arousal-based theory of personality, a number of problems refused to be silenced. For one, introverts show *weaker* classical conditioning under conditions conducive to high arousal (which, we must assume, is also induced by aversive UCSs), as seen in eyeblink conditioning studies (Eysenck and Levey, 1967). This finding supports Eysenck's *own* theory that introverts are transmarginally inhibited by high arousal, but *at the very same moment* fails to explain adequately the genesis of clinical neurosis. Other problems also screamed out to be heard. For example, impulsivity (inclined into the N plane; see Figure 11.1), not sociability (defining the extraversion axis), is often found to be associated with conditioning effects (Eysenck and Levey, 1972), but this places high arousability, and thus high conditionability, along an axis that is orthogonal to the one which has its high pole in the neurotic-introvert quadrant where clinical neurosis is located. Thus, Eysenck's *own* theory seems unable to explain the development of anxiety in neurotic-introverts. Time-of-day effects further undermine the central postulates of Eysenck's personality theory of clinical neurosis (see Gray, 1981).

In addition to the above problems, Gray cited a further reason to prefer a non-conditioning explanation (Corr, 2008). Now, classical conditioning theory states that as a result of the conditioned stimulus (CS) and unconditioned stimulus (UCS) being systematically paired, the CS comes to take on many of the eliciting properties of the UCS. That is, when presented alone after conditioning, the CS produces a response (i.e. the conditioned response, CR) that resembles the unconditioned response (UCR) elicited by the UCS. However, the CR *does not* substitute for the UCR – in several important respects, the CR does not even resemble the UCR. For example, a pain UCS will elicit a wide variety of reactions (e.g. vocalization and behavioural excitement) which are quite different to those elicited by a CS *signalling* pain, which consists of a quite different set of behaviours (e.g. quietness and behavioural inhibition). We thus have a theory that does not seem fit for purpose: classical conditioning cannot explain the pathogenesis or phenomenology of neurosis, although it can explain how initially neutral stimuli (CSs) acquire the motivational power to elicit this state. Gray asked the crucial question: if classical conditioning does not account for the generation of the negative emotional state that characterises neurosis, then what does? His answer – based upon extensive animal research (e.g. behavioural, pharmacological, lesion, and electrical stimulation studies) – was an innate mechanism, namely the *behavioural inhibition system* (BIS; Gray, 1976, 1982).

Three systems of 'classic' RST

RST gradually developed over the years to include three major systems of emotion:

1. The *behavioural inhibition system* (BIS) was postulated to be sensitive to *conditioned* aversive stimuli (i.e. signals of both punishment and the omission/termination of reward) relating to Anx, but also to extreme novelty, high-intensity stimuli, and innate fear stimuli (e.g. snakes, blood), which are more related to fear.

In addition, two other systems were postulated:

2. The *fight/flight system* (FFS) was postulated to be sensitive to *unconditioned* aversive stimuli (i.e. innately painful stimuli), mediating the emotions of rage and panic. This system was related to the state of negative affect (NA) (associated with pain) and speculatively associated by Gray with Eysenck's trait of psychoticism.
3. The *behavioural approach system* (BAS) was postulated to be sensitive to *conditioned* appetitive stimuli, forming a positive feedback loop, activated by the presentation of stimuli associated with reward and the termination/omission of signals of punishment. This system was related to the state of positive affect (PA) and the trait of Imp.

The BIS was modelled on the detailed pattern of behavioural effects of classes of drugs known to affect anxiety in human beings. By this route, Gray argued, anxiety could be operationally specified as those behaviours changed by anxiolytic drugs. Of course, there exists here the danger of circularity of argument; this was avoided by the postulation that anxiolytic drugs do not simply reduce anxiety (itself a vacuous tautology), but could be shown to have a number of behavioural effects in typical animal learning paradigms. Experimental evidence showed that anti-anxiety drugs affected responses to conditioned aversive stimuli, the omission of expected reward and conditioned frustration, all of which Gray postulated were mediated by a BIS, which was responsible for suppressing ongoing operant behaviour in the face of threat, as well as enhancing information processing and vigilance. (We shall see that in this revised theory, these effects can be reclassified as *conflict* effects.) Later, the BAS was added to account for behavioural reactions to rewarding stimuli – these were largely unaffected by anti-anxiety drugs. The danger of a circularity of argument was further reduced by the behavioural profile of the newer classes of anxiolytics which, it turned out, had the same behavioural effects and acted on the same neural systems as the older class of drugs, despite the fact that they had different psychopharmacological modes of action and side-effects (Gray and McNaughton, 2000).

Revised (2000–) RST

The Gray and McNaughton (2000) revised theory updates and extends the 'classic' version. These changes are, in parts, substantial: but, in other parts, more a clarification of the 1982 theory. Revised RST postulates three systems.

1. The *fight–flight–freeze system* (FFFS) is responsible for mediating reactions to aversive stimuli of all kinds, conditioned *and* unconditioned. It further proposes that there exists a hierarchical array of neural modules, responsible for avoidance and escape behaviours. Now, the FFFS mediates the emotion of fear, not anxiety. The associated personality factor comprises fear-proneness and avoidance, which is clinically mapped onto such disorders as phobia and panic.
2. The BAS mediates reactions to *all* appetitive stimuli, conditioned and unconditioned. This system generates the appetitively hopeful emotion of 'anticipatory pleasure', and hope itself. The associated personality comprises optimism, reward-orientation and impulsiveness, which clinically maps onto addictive behaviours (e.g. pathological gambling) and various varieties of high-risk, impulsive behaviour, and possibly the appetitive component of mania. The BAS is largely unchanged in the revised Gray and McNaughton version of RST.

3 The BIS is responsible, not, as in the 1982 version, for mediating reactions to conditioned aversive stimuli and the special class of innate fear stimuli, but for the resolution of *goal conflict* in general (e.g. between BAS-approach and FFFS-avoidance, as in foraging situations – but it is also involved in BAS–BAS and FFFS–FFFS conflicts). The BIS generates the emotion of anxiety, which entails the inhibition of prepotent conflicting behaviours, the engagement of risk assessment processes, and the scanning of memory and the environment to help resolve concurrent goal conflict.

The BIS resolves conflicts by increasing, through recursive loops, the negative valence of stimuli (these are adequate inputs into the FFFS), until behavioural resolution occurs in favour of approach or avoidance. Subjectively, this state is experienced as worry and rumination. The associated personality comprises worry-proneness and anxious rumination, leading to being constantly on the look-out for possible signs of danger, which map clinically onto such conditions as generalized anxiety and obsessional-compulsive disorder (OCD). There is an optimal level of BIS activation: too little leads to risk seeking (e.g. psychopathy) and too much to risk aversion (generalized anxiety), both reflecting suboptimal conflict resolution.

NEUROPSYCHOLOGICAL STRUCTURE OF THE REVISED THEORY

Revised RST agrees with the classical version in its assertion that substantive affective events fall into just two distinct major classes: positive and negative (Gray, 1975; Gray, 1982; Gray and McNaughton, 2000). Rewards and punishments are the obvious exemplars of positive and negative events, respectively. But, importantly for human experiments, the absence of an expected positive event is functionally the same as the presence of a negative event and vice-versa (Gray, 1975). Omission of expected reward is thus punishing. Similarly, the absence of an expected negative event is functionally the same as the presence of a positive event. Omission of punishment is rewarding. This basic scheme gives rise to a two-dimensional model of the neuropsychology of emotion, motivation, and personality that simplifies the theory, as well as serving as a point of unification of the otherwise complex arrangement of the separate neural modules underlying behaviour (McNaughton and Corr, 2004).

Fear and anxiety – defensive direction

The first dimension, 'defensive direction', is categorical. It rests on a functional distinction between behaviours that remove an animal from a source of danger (FFFS-mediated) and those that allow it cautiously to approach a source of potential danger (BIS-mediated). These functions are ethologically and pharmacologically distinct and, on each of these separate grounds, can be identified with fear and anxiety, respectively. The revised theory treats fear and anxiety as not only quite distinct but also, in a sense, as opposites. The categorical separation of fear from anxiety as classes of defensive responses has been demonstrated by Robert and Caroline Blanchard (Blanchard and Blanchard, 1988, 1990; Blanchard et al., 1997).

The Blanchards used 'ethoexperimental analysis' of the innate reactions of rats to cats to determine the functions of specific classes of behaviour. One class of behaviours was elicited by the immediate presence of a predator. This class could clearly be attributed to a state of fear. The behaviours, grouped into the class on purely ethological grounds, were sensitive to panicolytic drugs but not to drugs that are specifically anxiolytic. This is consistent with the insensitivity to anxiolytic drugs of active avoidance in a wide variety of species, and phobia in humans is also insensitive to anxiolytic drug treatment (Sartory et al., 1990). A second, quite distinct, class of behaviours (including 'risk assessment') was elicited by the potential presence of a predator.

This class of behaviours was sensitive to anxiolytic drugs. Both functionally and pharmacologically, this class was distinct from the behaviours attributed to fear and could be attributed to a state of anxiety.

Fear and anxiety – defensive distance

The second dimension, 'defensive distance', is graded: it rests on a functional hierarchy that determines appropriate behaviour in relation to defensive distance (i.e. perceived distance from threat). This second dimension applies equally to fear and anxiety but is instantiated separately in each.

Defensive distance equates with real distance; but in a more dangerous situation, the perceived defensive distance is shortened. In other words, defensive behaviour (e.g. active avoidance) will be elicited at a longer (objective) distance with a highly dangerous stimulus (which shortens *perceived* defensive distance), as compared to the elicitation of defensive behaviour by a less dangerous stimulus. According to the theory, certain individuals have a much shorter perceived defensive distance for a given threat stimulus, and thus react more intensively to relatively innocuous (in real distance terms) stimuli.

McNaughton and Corr (2004) view individual differences in defensive distance for a fixed real distance as a reflection of the personality dimension underlying 'punishment sensitivity', or 'threat perception'. They suggest that the high pole of this dimension is neurotic-introversion and the low pole is stable-extraversion. This personality dimension affects the FFFS-mediated behaviours directly, but affects those mediated by the BIS only indirectly (e.g. via FFFS-BAS goal conflict). Anxiolytic drugs are argued to alter (internally perceived) defensive distance relative to actual external threat. They *do not* affect defensive behaviour directly, but rather operate to shift behaviour along the defensive axis, often leading to the output of a different behaviour (e.g. risk-assessment to pre-threat behaviour).

An important conclusion of this theory, which goes to show the subtlety of revised RST, is the claim that the comparison of individuals on a single measure of performance at only a single level of threat may produce results that are difficult to interpret. For example, for an objectively defined defensive distance, one person may be in a state of panic and so cease moving, while another may actively avoid and so increase their movement. That is, highly sensitive and insensitive fearful individuals will show *different* behaviours *at the same level of threat* (defined in objective terms), as indeed will trait-identical individuals at different levels of threat. Thus moving people along this axis of defensive distance (by drugs or by experimental means) will not simply affect the strength or probability of a given behaviour, but is expected to result in different behaviours (which, themselves, may be in opposition). As we can see, at the core of the revised theory are ethological factors, relating specific behaviours to specific threats and environmental conditions.

Conflict

Revised RST defines anxiety in terms of defensive approach. However, this notion contains something more fundamental about anxiety, namely, *conflict*. An animal approaches a threat only if there is some possibility of a positive outcome (e.g. food when foraging in an unsafe field). But threats are not the only sources of aversion and avoidance encountered. In principle, approach–approach and avoidance–avoidance conflicts also involve activation of the same system and have essentially the same effects as classic approach–avoidance. It turns out that the conditioned stimuli to which the unrevised version of the BIS was said to be sensitive are, according to this formulation, specific examples of conflict stimuli. Thus, the new BIS theory reclassifies conditioned stimuli and expands the type of stimuli processed by the BIS. All of these now fall under the common rubric of goal conflict. This reformulation also helps tidy-up the rag-bag of other eliciting stimuli of the BIS (i.e. innate stimuli and high-intensity noise): in their non-conflict form, they now belong with the FFFS.

NEURAL SYSTEMS OF FEAR AND ANXIETY

Revised RST combines a large number of brain structures ranging from the prefrontal cortex, at the highest level, to the periaqueductal grey, at the lowest level, assigning to each structure: (a) a specific place in the theory; (b) a specific fundamental class of function; and (c) a specific class of mental disorder (McNaughton and Corr, 2008). Thus, the most fundamental change to the old view of the BIS is that it is *distributed* among a number of neural structures.

General architecture

The concepts of defensive direction and defensive distance provide a two-dimensional schema within which all defensive behaviours can be described. The theory translates this two-dimensional psychological schema into a matching two-dimensional neurological one. In particular, the categorical distinction *between* defensive approach and defensive avoidance is translated into two distinct parallel streams of neural structures; and the dimension of defensive distance is translated into the levels of a hierarchy of structures *within* each of the parallel streams (Figure 11.2).

The neural mapping of defensive distance into the two hierarchies is rendered simple by two architectural features. First, smaller defensive distances map to more caudal, subcortical neural structures while larger defensive distances map to more rostral, cortical neural structures with intermediate structures arranged in caudo-rostral order in between. Second, this mapping occurs in a symmetrical fashion with matching structures located within each of the parallel streams (this often involves subdivisions, or nuclei, of the same named area).

THE BEHAVIOURAL APPROACH SYSTEM (BAS)

We now have an outline of the FFFS and the matching components of the BIS. Revised RST theory also has a central place for the BAS. It must be borne in mind that, although the BIS would be activated with the simultaneous activation of the FFFS and the BAS (e.g. in the case of approach–avoidance conflict), it remains the case that the BAS is conceptually distinct from both the BIS and the FFFS.

Neural organization of the BAS

There are tensions in attempts to map the BAS onto brain systems and functions. As with the BIS and the FFFS, the BAS can be viewed as hierarchically organized. Gray (Gray and McNaughton, 1996; Gray et al., 1991) has described the BAS as having a 'caudate' component and an 'accumbens' component. However, he also made clear that 'accumbens holds a list of subgoals making up a given motor program and is able to switch through the list in an appropriate order, but to retrieve the specific content of each step, it needs to call up the appropriate subroutine by way of its connections to the [caudate] system' (Gray and McNaughton, 1996). Such caudate motor command subroutines are quite distinct from the affect-laden goals that are the subject of the FFFS, BAS and BIS (Gray and McNaughton, 2000).

On the other hand, as with the FFFS, the hierarchical organization of the BAS makes

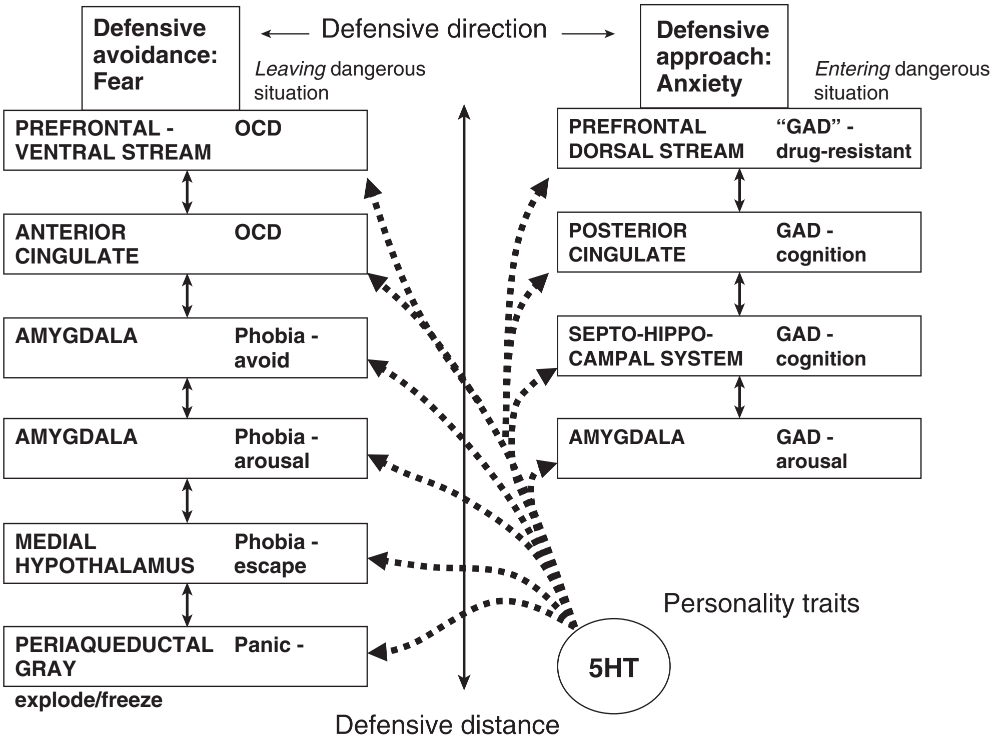

Figure 11.2 The two-dimensional defence system of fear and anxiety. On either side are defensive avoidance and defensive approach, respectively (this is a categorical dimension of 'defensive direction'). Each system is divided into a number of hierarchical levels (corresponding to the second dimension of 'defensive distance'). These are ordered from high to low (top to bottom) both with respect to neural level (and cytoarchitectonic complexity) and to functional level. Each level is associated with specific classes of behaviour and so symptom and syndrome (as shown). General monoamine modulation is shown as the putative 'personality' influence that provides unity to each system

it difficult for any part of it to control overall BAS sensitivity. Where a personality factor is thought to alter such sensitivity generally, we should probably look for appropriate modulatory systems. The neuromodulator that is probably of primary importance in BAS functioning is dopamine (DA; Depue and Collins, 1999; Pickering and Gray, 1999). The accumbens and caudate separation, alluded to by Gray, is reflected in the distinction between the so-called mesolimbic and nigrostriatal projection pathways of dopaminergic cells (these project to accumbens and caudate respectively, along with other structures). However, many influences (e.g. genes), which could generate individual differences in dopaminergic neurotransmission, may well express their effects on more than one dopaminergic projection system (Depue and Collins, 1999). Moreover, the structures innervated by these distinct dopaminergic systems act cooperatively to deliver behavioural responses thought of as being under BAS control.

In the neuroscience literature, over the last 15 years or so, a strong consensus has emerged over the functional significance of firing of dopaminergic cells in the midbrain

(Arbuthnott and Wickens, 2007; Schultz, 1998). The view is that DA cell firing reflects a 'reward prediction error' (RPE) signal. Specifically, in primates, increased bursts of DA cell firing result when an unexpected (under-predicted) reward occurs. Decreases in DA cell firing are observed when an expected reward does not occur (see Schultz, 1998, for details). Neuroimaging evidence in humans has also emerged which is consistent with this view (e.g. Abler et al., 2006). As argued elsewhere (Pickering and Gray, 1999, 2001; Pickering and Smillie, 2008), a proper neuroscientific understanding of the BAS will need to incorporate this RPE conceptualization of DA cell firing.

Of great interest in this area, the RPE view of DA cell firing is consistent with classic computational models of reinforcement learning (e.g. Dayan and Abbott, 2001). Learning in these models is hypothetically controlled by an RPE signal: a positive RPE (caused by an unexpected reward) is used to strengthen learning in the neural pathways which generated the behaviour leading to the reward; a large negative RPE (caused by a non-occurring expected reward) is used to extinguish learning in the neural pathways which generated the behaviour leading to the reward. When the RPE is close to zero (i.e. the level of reward is accurately predicted), then little learning takes place. The observations that DA cells fire in a fashion closely resembling an RPE signal was seen as providing a neural validation of these models. Moreover, the dopaminergic projection pathways release dopamine at sites very close to synapses on the dendritic spines of caudate and accumbens cells; these synapses are at the terminals of cortical inputs to the striatum. This synaptic arrangement, and the dendritic spines themselves, have a number of neurophysiological features (Wickens and Kotter, 1995) which enables an incoming burst of dopaminergic firing to operate effectively as a reinforcement/RPE signal and control learning at those cortico-striatal synapses.

The RPE conceptualization of dopamine cell firing in projections to BAS structures (caudate, accumbens, etc.) has strong resonances with the Gray and Smith (1969) cybernetic model of the functional interactions between the reward and punishment systems. In this model, the reward system had a comparator within it which determined whether the level of reward received matched the level expected. It was proposed that the results of this comparison process were fed back appropriately as inputs to the reward and punishment systems, although the detailed way in which this controlled learning of responses was not specified. The RPE account outlined above suggests how this learning may be accomplished. The Gray and Smith (1969) model proposed a general framework for choosing between responses leading to rewarding versus punishing behavioural consequences. Recent theoretical models of potential BAS structures in the basal ganglia have formalized the way that they may allow efficient decision-making of this kind (for an overview and references, see Bogacz, 2007).

Previously, accounts have been offered to begin to incorporate the neuroscience of dopamine cells and the basal ganglia into our understanding of the BAS (Pickering and Gray, 1999, 2001; see also Pickering and Smillie, 2008). This research is proceeding apace, and the final details have yet to be worked out. A challenge will be able to find an appropriate level of modelling which is able to distinguish between alternative neurally based accounts of the BAS.

What personality trait is linked to the BAS?

What broad personality trait might correspond to variations in the functioning of the BAS? Gray's original decision to call it 'impulsivity' was entirely ad hoc, as he repeatedly admitted. He used the ancient circular model of the humours (popularized by Eysenck) and drew a line between the types 'anxious' and 'carefree' (being confident that the BIS subserved trait anxiety). The line at right angles to the anxiety dimension (he assumed the BAS and BIS traits were orthogonal) approximately joins the labels of 'impulsive' and 'thoughtful' (although he might as easily have chosen 'optimistic' and 'careful' on these geometric grounds!). Thus, the impulsivity dimension was born; although Gray also had to decide which way round to place the dimension (high BAS types were assigned to the impulsive end of the dimension, on grounds of plausibility). This decision was further reinforced by the two components of extraversion in Eysenck's model, namely sociability and impulsivity, as well as experimental work showing impulsivity related to classical conditioning effects (see above).

On a related matter, Corr (2008) has drawn attention to the inadequate conceptualization of the BAS, especially as it relates to impulsivity. On evolutionary grounds, the BAS may be thought to be more complex than the FFFS, or indeed the BIS. The *primary* function of the BAS is to move the animal up the temporo-spatial gradient to the final biological reinforcer. This primary function is supported by a number of *secondary* processes, comprising perhaps simple approach, perhaps with BIS activation exerting behavioural caution at critical points, designed to reduce the distance between current and desired appetitive state (e.g. as seen in foraging behaviour in a densely vegetated field). However, in human behaviour, this depiction of BAS-controlled approach behaviour may be oversimplified.

First, it is helpful to distinguish the *incentive* motivation component and the *consummatory* component of reactions to appetitive stimuli. The neural machinery controlling reactions to unconditioned (innate) stimuli, and its associated emotion, must be different from that controlling the behaviour and emotion associated with *approach*, signalled by conditioned stimuli, to such stimuli. Thus, while the BAS responds to all appetitive stimuli, it is concerned specifically with the appetitive-approach aspects that move the animals towards the final biological reinforcer; at this point, non-BAS consummatory mechanisms, specific to the particular reinforcer concerned, are activated, e.g., the eating of food.

Second, moving to approach proper, we can discern a number of relatively separate, albeit overlapping, processes. At the simplest level, there seems an obvious difference between the 'interest' and 'drive' that characterizes the early stages of approach, and the behavioural and emotional excitement as the animal reaches the final biological reinforcer. Emotion in the former case may be termed 'anticipatory pleasure' (or 'hope'); in the latter, 'excitement'. There is evidence that, at the psychometric level, the BAS is multidimensional. For example, the Carver and White (1994) BIS/BAS scales measure three aspects of BAS: *reward responsiveness*, *drive,* and *fun-seeking*. It may be speculated that *drive* is concerned with actively pursing desired goals, *reward-responsiveness* is concerned with excitement at doing things well and winning, and *fun-seeking* is concerned with the impulsivity aspect of the BAS (which is especially appropriate for the capture of the final biological reinforcer).

Subgoal scaffolding

As discussed in detail by Corr (2008), BAS behaviour may best be seen as involving a series of appetitively motivated subgoals. That is, in order to move along the temporo-spatial gradient to the final primary biological reinforcer, it is necessary to engage in *subgoal scaffolding*. This process has several stages: (a) identification of the biological reinforcer; (b) planning behaviour; and (c) executing the plan. Important in this regard is the following: complex approach behaviour entails a series of behavioural processes, some of which oppose each other. For example, behaviour *restraint* and *planning* are often demanded to achieve BAS goals, but not at the final point of *capture* of the biological reinforcer, where non-planning and fast reactions (i.e. impulsivity) are more appropriate. Being a highly impulsive person – that is, acting fast without thinking and not planning – would not be appropriate BAS behaviour in anything other than very simple situations. Indeed, such behaviour would often move the animal *away* from their desired goal. For this reason, and others mentioned above, 'impulsivity' is not the most appropriate term for the personality factor corresponding to the full range of processes entailed by the BAS.

Therefore, given such a weak basis for Gray's initial labelling of the BAS, as well its apparent complexity, it is somewhat surprising that the BAS has been equated with impulsivity for so long. The first serious contradictory views came many years later. Depue and Collins (1999) argued that extraversion (and in particular its agentic aspects) better captured the nature of the BAS-related personality trait. Their argument drew on detailed support from the animal neurophysiological literature but was, in essence, a simple one. First, they suggested that the BAS was closely linked to dopaminergic neurotransmission. Second, they argued that the extant evidence pointed to a link between extraversion and dopaminergic neurotransmission which was stronger than the link for any other major personality trait. We (Corr, 1999; Pickering, 1999; Pickering and Gray, 1999) cautioned that the evidential basis for part two of their argument rested on a tiny body of data, mostly from Depue and colleagues' own laboratory. In addition, we suggested that Depue and Collins had ignored an equally small body of data which pointed to links between dopaminergic neurotransmission and a cluster of traits we have termed impulsive antisocial sensation seeking (ImpASS), rather than extraversion. At that time we felt that the jury could not reach as clear a verdict as that reached by Depue and Collins and argued that (aspects of) the ImpASS trait cluster might correspond to the BAS trait. Subsequent neuroscience data has emerged that is broadly in line with Depue and Collins' thesis (e.g. Cohen et al., 2005; Wacker et al., 2006). However, there are also psychometric and behavioural data (see Smillie et al., 2006, for a review) which we feel now tip the scales more strongly in favour of the idea that extraversion might be the BAS trait. But, further data are needed, especially ones relating specific psychometric measures of the revised RST's systems to extraversion.

INTERACTIONS OF THE BAS, FFFS, AND BIS: IMPLICATIONS FOR TRAIT MEASUREMENT

The old description of RST supposed that each system had a reactivity/sensitivity to its key inputs, which we can denote w_A, w_I, and w_F for the sensitivity of the BAS, BIS, and FFFS, respectively. Interindividual variations in w_A, w_I, and w_F are assumed to follow a normal distribution with each sensitivity independent of (uncorrelated with) the others. The trait of anxiety, Anx, was taken to reflect variation in w_I and another trait ('the BAS trait') was taken to reflect variation in w_A.

Elsewhere we (Corr, 2002; Pickering, 1997) argued that the effects of such systems on behaviour would generally not be independent of one another even though the sensitivities were themselves independent – although, under certain conditions, they would (specified by Corr, 2002). Thus, for example, a behaviour controlled by reward reinforcers would not only be influenced by the BAS personality trait (i.e. w_A) but could also often be influenced by Anx. Corr (2002) dubbed this the *joint subsystems hypothesis* in contrast to an earlier view that behaviour controlled by reward would depend selectively upon w_A (the *separable subsystems hypothesis*).

Recently, Smillie et al. (2006) took this view further. They argued that self-report questionnaire responses, used to measure personality traits, are likely to reflect subjective estimates of the functional *outcomes* rather than latent properties of the individual neural systems. A functional outcome of the BAS might be its mean output level across a range of situations, whereas a latent property would be its sensitivity (w_A). They suggested that the functional outcome will be available for introspection (and hence self-report) whereas a sensitivity will not, although the sensitivities will clearly have a direct influence on the observable functional outcome (someone with a higher value of w_A will, all other things being equal, have a higher mean BAS output level than a person with lower w_A). Looking at the item content of various possible BAS personality trait measures, Pickering (2008) concluded that such questionnaires might well reflect functional BAS outcomes (such as mean output level).

This viewpoint leads to some potentially striking conclusions. The functional outcomes of each system are, as for other reinforcer-controlled behaviours, likely to be susceptible to the joint influences of the various interacting systems. Smillie et al. (2006) report the results of simulation studies which illustrate this point. For one particular plausible set of interactions between the BIS, BAS and FFFS (in line with the revised Gray and McNaughton, 2000, model) they simulated functional outcomes (in this case mean output) across 200 randomly sampled and widely varying combinations of reinforcers. The mean BAS output across simulated individuals was predicted ($R^2 = 0.89$) by the following regression equation:

$$\textit{Mean BAS output} = (\beta_A \times w_A) - (\beta_F \times w_F) - (\beta_I \times w_I)$$

where the βs are positive-valued regression coefficients. The same model showed that mean BIS output was predicted ($R^2 = 0.85$) by:

$$(\beta'_A \times w_A) + (\beta'_F \times w_F) + (\beta'_I \times w_I)$$

By contrast, it is interesting to note that the mean FFFS output was predicted ($R^2 = 0.82$) only by the sensitivity of the FFFS.

Assuming some trait questionnaires do reflect functional outcomes of specific systems then these simulations raise important and paradoxical results. For example, the 'BAS-related' trait measures is BAS-related because it is defined by the functional outcome of the BAS and yet it is influenced by the sensitivities of all three interacting systems (w_A, w_F and w_I). Thus, if one were to develop a new BAS trait measure then one should not consider it invalidated if it correlated negatively with anxiety (BIS trait) measures; the simulations predict that such trait correlations should be observed. These predictions occur, it is worth reiterating, even though w_A and w_I (the underlying system sensitivities) are independent of one another. The description of the 'reinforcement sensitivity' theory of personality has implied a one-to-one mapping of traits (e.g. anxiety) onto the sensitivities of single systems (e.g. the BIS). The simulations show that this need not be the case and trait measures may be jointly determined by the sensitivities of all three interacting systems. It remains sensible, however, to talk of the theory as 'reinforcement sensitivity' theory, as the resulting personality traits are determined by the sensitivities of reinforcement-dependent systems; however, the one-to-one mapping of traits onto sensitivities is now being questioned.

In a speculative footnote to this section, we consider whether there might be some trait measures which line up more directly with underlying sensitivities rather than functional outcomes? The simulations suggested that, for traits related to FFFS functioning, the two bases (sensitivities, functional outcomes) may sometimes be more or less interchangeable. This fits well with the account proposed by McNaughton and Corr (2004, 2008) in which the trait of fearfulness (neurotic-introversion to stable-introversion) maps directly onto underlying punishment sensitivity.

However, one might also imagine a situation in which a trait measure, *T*, had items which reflected the functional outcome of one system along with other items which reflected the functional outcome of another system (we finesse here the question of whether such a trait measure could ever emerge in a factor analytic approach to trait measure development). Imagine such a measure was based on a mixture of BAS and BIS functional outcomes. The final trait measure (from the results of the simulations presented earlier) would be given by a summation of the two earlier regression equations:

$$T = (\beta_A \times w_A) - (\beta_F \times w_F) - (\beta_I \times w_I) + (\beta'_A \times w_A) + (\beta'_F \times w_F) + (\beta'_I \times w_I)$$

Assuming the values of β_I and β'_I, and β_F and β'_F, were broadly similar then the above equation would approximately reduce to

$$T = (\beta_A + \beta'_A) \times w_A$$

In this scenario, the trait measure *T* would directly reflect the sensitivity of a single underlying system (the BAS in this example).

High scores on such a trait measure would be found in people who had higher BAS functional outcomes (e.g. higher mean BAS outputs) *and* higher BIS functional outcomes (e.g. higher mean BIS outputs) across a range of situations. Is such a trait measure likely? Do any existing trait measures plausibly satisfy such conditions? We do not think this is likely. It might be suggested that extraversion questionnaires might be candidates for traits like *T* above. The EPQ extraversion scale, for example, has several items about enjoying social situations (e.g. Do you enjoy meeting new people? Would you enjoy yourself at a lively party?); these can plausibly be viewed by indexing mean BAS output in these contexts. However, under Gray and McNaughton's (2000) reformulation of RST, and based on the description of the action of the BIS, someone with a high mean BIS output would often be rather cautious and deliberate, tending to seek extra information when situations are ambiguous or when motivations are conflicting, and so on. Such a person might be described as low impulsive and deliberate. Items addressing these behavioural aspects might be found on some extraversion scales, and items addressing these behaviours on other scales would be very likely to correlate moderately with traditional extraversion items. However the correlation would be the opposite way round to that required for a trait measure such as *T* above; in our view, a trait measure like *T* therefore seems very unlikely to exist.

In summary, the main message of this section remains: the role of underlying reinforcement sensitivities in our revised understanding of RST seems likely to be more complex than has been hitherto suggested. With the possible exception of the punishment/fear system, variations in the sensitivities of the underlying systems to their characteristic inputs may not have one-to-one mappings onto observable personality traits.

PERSONALITY AND PSYCHOPATHOLOGY

How does personality relate to psychological conditions (e.g. anxiety). No doubt, the details of RST shall continue to undergo continual refinement and change – that is in the nature of any scientific theory – but we believe that 'defensive distance' and 'defensive direction' shall continue to play a pivotal role as they map onto a series of distinct neural modules, to each of which can be attributed a particular class of function, and so generation of a particular symptomatology (e.g. panic, phobia, obsession). As noted by McNaughton and Corr (2004, 2008), these 'symptoms' may be generated in several different ways:

1 as a normally adaptive reaction to specific (mild) eliciting stimuli (e.g. mild anxiety just before an exam);
2 as excessive activation of a related structure by its specific (strong) eliciting stimuli, but where the 'symptoms' are not excessive given the level

of input from the related structure (e.g. panic when crossing a railway line at the sight of a rapidly oncoming train);

3 at maladaptive intensity, as a result of excessive sensitivity to their specific eliciting stimuli (e.g. fearful avoidance as a result of seeing a harmless spider) – this would be a pathological reaction.

In addition, pathologically excessive (BIS) anxiety could generate (FFFS) panic with the latter being entirely appropriate to the level of apprehension experienced. Conversely, pathological panic could, with repeated experience, condition anxiety with the level of the latter being appropriate to the panic experienced. This modular view of the defence system, separated into distinct syndrome and symptom-specific, components was developed largely on the basis of animal experiments. In addition, the linking of this view to terms such as panic, phobia, and obsession is also justified by the clinical effects of drugs when taken together as a class. (All drugs have common and unique effects, and it is only their common effects that interest us here.) RST may provide a satisfactory explanation of the variety of clinical 'neurotic' phenomena observed, yet at the same time, may appear to destroy the very unity of an underlying personality trait.

However, this problem seems worse than it is. For rescue, we need only appeal to the fact that, based on quantitative genetic studies, there is a common fundamental predisposition to the plethora of clinical neurotic conditions observed, even though that predisposition manifests differently in different individuals (Kendler et al., 2003). Indeed, the action of many clinically effective drugs is best viewed as an interaction with more global modulatory systems. For example, 5HT neurons innervate virtually the entire defence system; and drugs such as imipramine or specific serotonin reuptake inhibitors (SSRIs), have a general effect on 5HT synapses. Such drugs affect anxiety, depression and panic because they increase the levels of 5HT in the different parts of the system controlling each.

Therefore, comparison of drug classes can be used to dissect out different parts of the defence system. But this comparison must involve several different drugs within each class if specific conclusions are to be drawn about specific brain systems. Conversely, the systems as a joint whole, and each system individually, may be globally susceptible to modulation controlled by the biological substrates underlying personality. In detail, then, the system underlying clinical drug action consists of two sets of parallel, interconnected modules dealing with defensive avoidance and defensive approach, respectively. Superimposed on these specialized modules are general modulatory systems.

It should be expected that if these modulatory systems are crucial for personality, there is also a conceptual need for general control. Certainly with the BIS, anxiolytics clearly alter defensive distance: they alter at whatever point of the neural hierarchy is in control given progressive variations in the external situation, and they do so in a lawful manner. Assuming that the control of fear by the monoamines operates in a similar manner to the control of anxiety by anxiolytic drugs, we should expect the personality factor related directly to 'punishment sensitivity' would be the one that alters the internal defensive distance in relation to any particular real distance. Put another way, a personality factor of fearfulness multiplies the quantum of fear inherent in a particular stimulus, producing many different levels (across different individuals) with the same stimulus.

CONCLUSIONS

There remains some considerable uncertainty as the best way to relate fundamental systems of emotion and motivation to personality factors, yet we contend that considerable progress has already been made. This chapter has illustrated that there is a lot of new theorizing which has substantially reformulated a popular theory of personality. As yet, however, this new thinking has not stimulated many new empirical findings. We hope that this situation will change in the near future. In relation to this issue, Smillie et al. (2006: 320) note that although RST is most often seen as a theory of anxiety and impulsivity, it is 'more accurately identified as a neuropsychology of emotion, motivation and learning. In fact, RST was born of basic animal learning research, initially not at all concerned with personality.' They go on to remark, 'RST did not develop as a theory *of* specific traits, but as a theory of specific biological systems which were later suggested to relate, *inter alia*, to personality' (2006: 321).

There is a related reason why basic emotion and motivation systems do not map neatly onto personality factors: basic emotion and motivation theory has extended beyond the point at which Gray suggested that the BIS and BAS relate to anxiety and impulsivity, respectively. Furthermore, RST personality researchers have developed scales to measure the BIS and BAS that were influenced by Gray's original thinking but which do not reflect more recent developments in the basic theory. Thus, RST research represents two distinct bodies of knowledge, the first concerned with neural systems and processes, the second with personality and its measurement. One of our purposes in writing this chapter is to encourage other researchers to work to bring these two aspects into closer alignment. Nonetheless, the Janus-faced nature of RST has also been a strength, making it a dynamically evolving theory, but it also poses obvious problems for, at any given time, specifying a consensual model agreed by researchers.

REFERENCES

Abler, B., Walter, H., Erk, S., Kammerer, H. and Spitzer, M. (2006) 'Prediction error as a linear function of reward probability is coded in human nucleus accumbens', *NeuroImage*, 31(2): 790–5.

Arbuthnott, G.W. and Wickens, J. (2007) 'Space, time and dopamine', *Trends in Neurosciences*, 30(2): 62–9.

Blanchard, D.C. and Blanchard, R.J. (1988) 'Ethoexperimental approaches to the biology of emotion', *Annual Review of Psychology*, 39: 43–68.

Blanchard, R.J. and Blanchard, D.C. (1990) 'An ethoexperimental analysis of defense, fear and anxiety', in N. McNaughton and G. Andrews (eds), *Anxiety*. Dunedin: Otago University Press, pp. 12–133.

Blanchard, R.J., Griebel, G., Henrie, J.A. and Blanchard, D.C. (1997) 'Differentiation of anxiolytic and panicolytic drugs by effects on rat and mouse defense test batteries', *Neuroscience and Biobehavioral Reviews*, 21(6): 783–9.

Bogacz, R. (2007) 'Optimal decision-making theories: Linking neurobiology with behaviour', *Trends in Cognitive Sciences*, 11(3): 118–25.

Carver, C.S. and White, T.L. (1994) 'Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales', *Journal of Personality and Social Psychology*, 67(2): 319–33.

Cohen, M.X., Young, J., Baek, J.M., Kessler, C. and Ranganath, C. (2005) 'Individual differences in extraversion and dopamine genetics reflect reactivity of neural reward circuitry', *Cognitive Brain Research*, 25(3): 851–61.

Corr, P.J. (1999) 'Does extraversion predict positive incentive motivation?', *Behavioral and Brain Sciences*, 22(3): 520–1.

Corr, P.J. (2001) 'Testing problems in J.A. Gray's personality theory: A commentary on Matthews and Gilliland (1999)', *Personal Individual Differences*, 30(2): 333–52.

Corr, P.J. (2002) 'J.A. Gray's reinforcement sensitivity theory: Tests of the joint subsystem hypothesis of anxiety and impulsivity', *Personality and Individual Differences*, 33(4): 511–32.

Corr, P.J. (2004) 'Reinforcement sensitivity theory and personality', *Neuroscience and Biobehavioral Reviews*, 28(3): 317–32.

Corr, P.J. (2008) 'Reinforcement sensitivity theory (RST): Introduction', in P.J. Corr (ed.)

The Reinforcement Sensitivity Theory of Personality. Cambridge: Cambridge University Press, pp. 1–43.

Corr, P.J. and McNaughton, N. (2008). 'Reinforcement sensitivity theory and personality', in P.J. Corr (ed.), *The Reinforcement Sensitivity Theory of Personality.* Cambridge: Cambridge University Press, pp. 155–87.

Dayan, P. and Abbott, L.F. (2001) Theoretical Neuroscience: *Computational and Mathematical Modeling of Neural Systems.* Cambridge, MA: MIT Press.

Depue, R.A. and Collins, P.F. (1999) 'Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion', *Behavioral and Brain Sciences*, 22(3): 491–517.

Eysenck, H.J. (1944) 'Types of personality: A factorial study of 700 neurotics', *Journal of Mental Science*, 90: 859–61.

Eysenck, H.J. (1947) *Dimensions of Personality*. London: K. Paul, Trench Trubner.

Eysenck, H.J. (1957) *The Dynamics of Anxiety and Hysteria*. New York: Preger.

Eysenck, H.J. (1967) *The Biological Basis of Personality*. Springfield, IL: Thomas.

Eysenck, H.J. (1979) 'The conditioning model of neurosis', *Behavioural and Brain Sciences*, 2(2): 155–99.

Eysenck, H.J. and Eysenck, M.W. (1985) *Personality and Individual Differences: A Natural Science Approach*. New York: Plenum Press.

Eysenck, H.J. and Levey, A. (1972) 'Conditioning, introversion–extraversion and the strength of the nervous system', in V.D. Nebylitsyn and J.A. Gray (eds), *The Biological Bases of Individual Behaviour*. London: Academic Press. pp. 206–20.

Gray, J.A. (1970) 'The psychophysiological basis of introversion–extraversion', *Behaviour Research and Therapy*, 8(3): 249–66.

Gray, J.A. (1972a) 'Learning theory, the conceptual nervous system and personality', in V.D. Nebylitsyn and J.A. Gray (eds), *The Biological Bases of Individual Behaviour*. New York: Academic Press. pp. 372–99

Gray, J.A. (1972b) 'The psychophysiological nature of introversion–extraversion: A modification of Eysenck's theory', in V.D. Nebylitsyn and J.A. Gray (eds), *The Biological Bases of Individual Behaviour*. New York: Academic Press. pp. 182–205.

Gray, J.A. (1975) *Elements of a Two-Process Theory of Learning*. London: Academic Press.

Gray, J.A. (1976) 'The behavioural inhibition system: A possible substrate for anxiety', in M.P. Feldman and A.M. Broadhurst (eds), *Theoretical and Experimental Bases of Behaviour Modification*. London: Wiley. pp. 3–41.

Gray, J.A. (1981) 'A critique of Eysenck's theory of personality', in H.J. Eysenck (ed.), *A Model for Personality*. Berlin: Springer. pp. 246–76.

Gray, J.A. (1982) *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Oxford: Oxford University Press.

Gray, J.A., Feldon, J., Rawlins, J.N.P., Hemsley, D.R. and Smith, A.D. (1991) 'The neuropsychology of schizophrenia', *Behavioral and Brain Sciences*, 14(1): 1–84.

Gray, J.A. and McNaughton, N. (1996) 'The neuropsychology of anxiety: Reprise', in D.A. Hope (ed.), *Perspectives on Anxiety, Panic and Fear*. Nebraska: University of Nebraska Press. pp. 61–134.

Gray, J.A. and McNaughton, N. (2000) *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Oxford: Oxford University Press.

Gray, J.A. and Smith, P.T. (1969) 'An arousal decision model for partial reinforcement and discrimination learning', in R.M. Gilbert and N.S. Sutherland (eds), *Animal Discrimination Learning*. London: Academic Press. pp. 243–72.

Hebb, D.O. (1955) 'Drives and the C.N.S. (Conceptual Nervous System)', *Psychological Review*, 62(4): 243–54.

Kendler, K.S., Prescott, C.A., Myers, J. and Neale, M.C. (2003) 'The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women', *Archives of General Psychiatry*, 60(9): 929–37.

McNaughton, N. and Corr, P.J. (2004). 'A two-dimensional neuropsychology of defense: Fear/anxiety and defensive distance', *Neuroscience and Biobehavioral Reviews*, 28(3): 285–305.

McNaughton, N. and Corr, P.J. (2008). 'The neuropsychology of fear and anxiety: A foundation for reinforcement sensitivity theory', in P.J. Corr (ed.), *The Reinforcement*

Sensitivity Theory of Personality. Cambridge: Cambridge University Press, pp. 44–94.

Pavlov, I.P. (1927) *Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford: Oxford University Press. (Translated and edited by G.V. Anrep.)

Pickering, A.D. (1997) 'The conceptual nervous system and personality: From Pavlov to neural networks', *European Psychologist*, 2(2): 139–63.

Pickering, A.D. (1999) 'Personality correlates of the dopaminergic facilitation of incentive motivation: Impulsive sensation seeking rather than extraversion?', *Behavioural and Brain Sciences*, 22(3): 534–5.

Pickering, A.D. (2008) 'Formal and computational models of reinforcement sensitivity theory', in P.J. Corr (ed.), *The Reinforcement Sensitivity Theory of Personality*. Cambridge: Cambridge University Press, pp. 453–81.

Pickering, A.D., Corr, P.J. and Gray, J.A. (1999) 'Interactions and reinforcement sensitivity theory: A theoretical analysis of Rusting and Larsen (1997)', *Personality and Individual Differences*, 26(2): 357–65.

Pickering, A.D. and Gray, J.A. (1999) 'The neuroscience of personality', in L. Pervin and O. John (eds), *Handbook of Personality* (2nd edition). New York: Guilford Press. pp. 277–99.

Pickering, A.D. and Gray, J.A. (2001) 'Dopamine, appetitive reinforcement, and the neuropsychology of human learning: An individual differences approach', in A. Eliasz and A. Angleitner (eds), *Advances in Individual Differences Research*. Lengerich, Germany: PABST Science Publishers. pp. 113–49

Pickering, A.D. and Smillie, L.D. (2008) 'The behavioural activation system: Challenges and opportunities', in P.J. Corr (ed.), *The Reinforcement Sensitivity Theory of Personality*. Cambridge: Cambridge University Press, pp. 120–54.

Sartory, G., MacDonald, R. and Gray, J.A. (1990) 'Effects of diazepam on approach, self-reported fear and psychophysological responses in snake phobics', *Behaviour Research and Therapy*, 28(4): 273–82.

Schultz, W. (1998) 'Predictive reward signal of dopamine neurons', *Journal of Neurophysiology*, 80(1): 1–27.

Smillie, L.D., Pickering, A.D. and Jackson, C.J. (2006) 'The new reinforcement sensitivity theory: Implications for personality measurement', *Personality and Social Psychology Review*, 10(4): 320–35.

Wacker, J., Chavanon, M. and Stemmler, G. (2006) 'Investigating the dopaminergic basis of extraversion in humans: A multilevel approach', *Journal of Personality and Social Psychology*, 91(1): 171–87.

Wickens, J. and Kotter, R. (1995) 'Cellular models of reinforcement', in J.C. Houk, J.L. Davis and D.G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*. London: MIT Press, pp. 189–214.