Personality and Social Psychology Review

http://psr.sagepub.com/

The New Reinforcement Sensitivity Theory: Implications for Personality Measurement

Luke D. Smillie, Alan D. Pickering and Chris J. Jackson

Pers Soc Psychol Rev 2006 10: 320
DOI: 10.1207/s15327957pspr1004_3

The online version of this article can be found at:
http://psr.sagepub.com/content/10/4/320

Published by:

SAGE

http://www.sagepublications.com

On behalf of:

Society for Personality and Social Psychology

Additional services and information for *Personality and Social Psychology Review* can be found at:

Email Alerts: http://psr.sagepub.com/cgi/alerts

Subscriptions: http://psr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://psr.sagepub.com/content/10/4/320.refs.html

>> Version of Record - Nov 1, 2006

What is This?

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

Personality and Social Psychology Review
2006, Vol. 10, No. 4, 320–335

Copyright © 2006 by
Lawrence Erlbaum Associates, Inc.

The New Reinforcement Sensitivity Theory: Implications for Personality Measurement

Luke D. Smillie
Alan D. Pickering
Department of Psychology
Goldsmiths College, University of London

Chris J. Jackson
School of Psychology
University of Queensland

In this article, we review recent modifications to Jeffrey Gray's (1973, 1991) reinforcement sensitivity theory (RST), and attempt to draw implications for psychometric measurement of personality traits. First, we consider Gray and McNaughton's (2000) functional revisions to the biobehavioral systems of RST. Second, we evaluate recent clarifications relating to interdependent effects that these systems may have on behavior, in addition to or in place of separable effects (e.g., Corr, 2001; Pickering, 1997). Finally, we consider ambiguities regarding the exact trait dimension to which Gray's "reward system" corresponds. From this review, we suggest that future work is needed to distinguish psychometric measures of (a) fear from anxiety and (b) reward-reactivity from trait impulsivity. We also suggest, on the basis of interdependent system views of RST and associated exploration using formal models, that traits that are based upon RST are likely to have substantial intercorrelations. Finally, we advise that more substantive work is required to define relevant constructs and behaviors in RST before we can be confident in our psychometric measures of them.

The development of a personality taxonomy typically begins with a descriptive measurement model, that is, with the identification and quantification of trait dimensions. Various methods are employed for identifying these traits, ranging from factor analyses of psychological adjectives (e.g., Cattell, 1950; Goldberg, 1982), to extrapolation of Jungian philosophy (Briggs & Myers, 1980). In some cases, the usefulness of the resultant descriptions and structural robustness of the taxonomy form the sole focus or "end-state" for the model (e.g., the various manifestations of the "Five-Factor" or "Big Five" approaches, as reviewed by Block, 1995). In other cases, once descriptive traits have been identified, underlying causes of individual variation on measures of these traits are then investigated (e.g., Eysenck, 1967; Zuckerman, 1979). Although this summary might not perfectly represent all approaches to the study of personality, it is intended to reflect the central focus of most theorists upon particular trait dimensions.

A unique exception to this picture is Jeffrey Gray's (1970, 1973, 1982b, 1987a) reinforcement sensitivity theory (RST). This account is perhaps best known as a theory of *Anxiety* and *Impulsivity*, although is more accurately identified as a neuropsychology of emotion, motivation, and learning. In fact, RST was born of basic animal learning research (as reviewed by Gray, 1987b), initially not at all concerned with personality. (To clarify, RST was primarily concerned with Anxiety, but more in terms of a spectrum of disorders and dispositions rather than a descriptive personality dimension.) It is from this field that the biological systems that have come to play the central role in Gray's theory were identified, and their functioning observed to vary between individuals (rodents) in a stable and heritable manner (see especially Gray, 1987b, pp. 41–50, and Gray & McNaughton, 2000, pp. 342–349). Then, with Gray's (1981) landmark critique of Hans Eysenck's (1967) arousal-activation model of personality, the way was paved for what has been described as a revolution in personality psychology (Depue & Collins, 1999): the proposal that motivation and emotion may comprise the central processes underlying trait di-

This article has benefited from the advice of several individuals, to whom we owe our sincere gratitude. We especially thank Neil McNaughton at the University of Otago for his helpful thoughts on this article as it developed. We also thank the two reviewers for their time and effort in providing us with detailed constructive feedback on an earlier version of this article.

Correspondence should be sent to Luke D. Smillie, Department of Psychology, Goldsmiths College, The University of London, New Cross SE14 6NW, UK. E-mail: l.smillie@gold.ac.uk

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

mensions. The significance of this shift in thinking has been remarked upon often enough to dissuade us from doing so here—our point is that RST did not develop as a theory of specific traits, but as a theory of specific biological systems that were suggested to relate, *inter alia*, to personality.

A consequence arising from this peculiarity of the development of RST is an apparent duality of the model. Basic research concerning emotion and motivation has continued beyond Gray's original (e.g., 1982b) observations, and independently from personality psychology. Simultaneously, Gray's views concerning Impulsivity and Anxiety, the two traits that RST offers an explanation, have proven attractive to personality researchers. This influence is visible in the appearance of many RST-based personality scales (e.g., Carver & White, 1994; McAndrew & Steel, 1991; Torrubia, Ávila, Moltó, & Caseras, 2001), as well as in the incorporation of Gray's postulates within other prominent personality models (e.g., Cloninger, 1987; Zuckerman, 1979). Overall then, it is reasonable to describe RST research as two distinct bodies of inquiry extending from a fork in the road: The first is concerned with the physiological hardware, conceptual organization, and functional nature of certain motivational systems, whereas the second is concerned with the theoretical understanding, psychometric measurement, and predictive utility of the traits Anxiety and Impulsivity.

These two faces of RST are not only distinct domains of research, but also divergent, to the point that they have become increasingly difficult to reconcile. We argue that this state of affairs has had particular impact upon the trait view of RST. As if it were frozen in time, Gray's "personality model" is a relatively discrete slice of an otherwise continuous and ongoing field of knowledge. This might not create problems for other personality models, where the ongoing research is largely in-the-service-of the trait taxonomy. For instance, the dimensions of the Big Five factor model are held to be "real, pervasive, universal and biologically based" (Costa & McCrae, 1992, p. 861), and, as such, most Big Five research aims to develop our understanding of that trait system. In comparison, RST research is in no way constrained by the descriptive trait system suggested by Gray. Explanation of personality is a compelling by-product of RST, but a by-product nonetheless. Indeed, of a seminal work to which most RST articles in the personality literature refer (Gray, 1982b, *The Neuropsychology of Anxiety*), Gray noted that "personality was only briefly touched upon in this book" (Gray, 1982a, p. 525). The almost indifferent attitude of RST to the trait system with which it is linked has been demonstrated by a number of major revisions to the model over the last decade, many of which contradict fundamental assumptions of the trait system (Corr, 2004). As might be expected of an unprepared audience, personality researchers have so far made little attempt to consider the potential implications of these changes for trait conceptualization and measurement. The purpose of this review, therefore, is to begin such an appraisal.

The Original Reinforcement Sensitivity Theory

The original version of RST (e.g., as summarized by Gray, 1987a, 1991), and the version to which most researchers still implicitly refer, consisted of a reward system, a punishment system, and a threat-response system. These were known as the Behavioral Inhibition System (BIS), the Behavioral Activation System (or Behavioral Approach System; BAS), and the Fight/Flight System (FFS). The BIS—which was the central focus of Gray's work—was thought to mediate responses to conditioned signals of punishment (resulting in passive-avoidance) and conditioned signals of frustrative nonreward (resulting in extinction of a response), and was suggested to provide the causal basis of Anxiety. The BAS was thought to mediate responses to conditioned signals of reward (resulting in approach behavior) and conditioned signals of relieving nonpunishment (resulting in active-avoidance), and was suggested to provide the causal basis of Impulsivity. Anxiety and Impulsivity are typically represented graphically as a 45° rotation to Eysenck's (1967) extraversion-neuroticism model (see Figure 1), although this was always acknowledged to be a simplification; the correct angle is thought to be around 30° (Pickering, Corr, & Gray, 1999).

In addition to behavioral and trait manifestations of the BAS and BIS, their activation was thought to give rise to the affective dimensions of positive and negative mood. Activation of the BAS by reward was thought to lead to increased positive affect, and activation of the BIS by punishment was thought to lead to increased negative affect. As a result, individual differences in BAS and BIS reactivity were thought to correspond to stable differences in positive and negative emotionality (Fowles, 1993; Gray, 1994). Although the implications of RST for affect and emotion are beyond the scope of this review (as we are principally concerned with implications for psychometric measurement of personality), it is worth noting briefly that those implications appear to be more complex than has been traditionally presumed. This is particularly the case for the BAS, which has been linked with a range of emotional reactions to negative stimuli (Carver, 2004). For example, Carver showed that a measure of BAS (Carver & White's BAS scale, 1994) predicts self-reported frustration in response to the omission of anticipated reward. This is inconsistent with the original version of RST but highly consistent with more recent neuro-

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

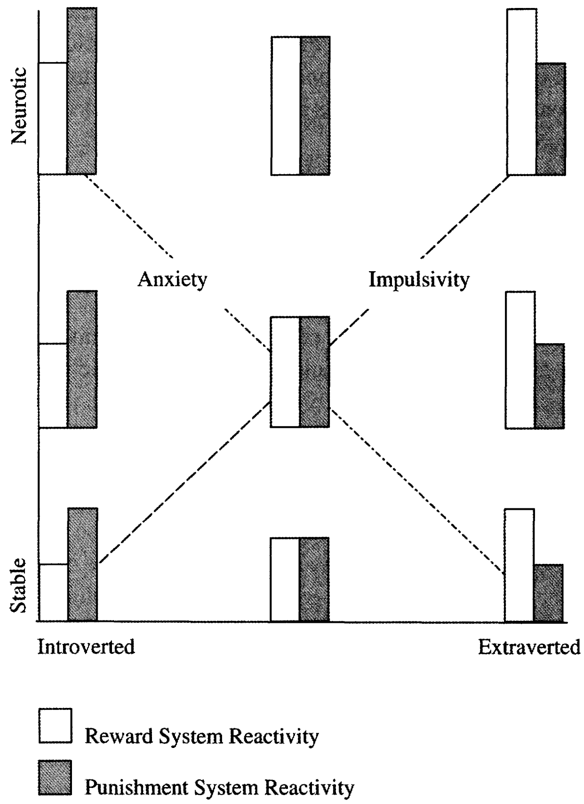

Figure 1. Diagrammatic representation of Gray's modification of Eysencks's (1967) trait system according to the alternative theoretical framework provided by reinforcement sensitivity theory (Pickering, Corr, & Gray, 1999). As one moves along the Impulsivity dimension, the amount of reward reactivity increases but punishment reactivity does not. Similarly, as one moves along the Anxiety dimension the amount of punishment reactivity increases but reward reactivity does not. Some earlier versions of this figure do not show accurate absolute magnitudes of reward/punishment system reactivities.

psychological models of reward-contingent learning (e.g., Frank, 2005; Schultz, 1998), in which the same neural signal (phasic dopamine firing) is responsive to both reward and nonreward events.

Gray's (1987b) remaining system, the FFS, was thought to mediate responses to unconditioned aversive stimuli, resulting in rapid escape (flight) or defensive aggression (fight). Situational factors were a key determinant of the output of this system, such that flight resulted if the threat was distal and/or able to be escaped, otherwise fight resulted. Fight and flight were respectively linked to the emotional states of rage and fear (Gray, 1987b), but were never clearly linked with personality (Corr, 2001; Jackson, 2003). As the FFS was essentially a secondary punishment system, and had a strikingly similar role to the BIS, the distinctiveness of these systems in the original version of RST was hazy. Flight was always distinguished from BIS-mediated avoidance as it was an unlearned explosive response, but the boundary between fear and anxiety was difficult to discern. As a result, equivalence has often been assumed between fear and anxiety, both conceptually and in terms of psychometric measurement (White & Depue, 1999).

Gray (1973) suggested that the FFS might be linked with Eysenck's dimension of Psychoticism (Eysenck & Eysenck, 1968), and this is indeed the trait with which many syntheses of personality have aligned the threat-response system (e.g., Revelle, 1995). However, there are substantive differences between behaviors that would be related to Psychoticism on the one hand, and the FFS on the other. Perhaps the most notable of these are anger and aggression, which are dispositionally affected by Psychoticism in Eysenckian theory (Eysenck & Gudjonsson, 1989) and might therefore be presumed to be related to Fight in RST. Fight is, however, an explosive and disorganized reaction to inescapable pain or intense threat, and is thus primarily an expression of fear. Therefore, it should not be confused with predatory aggression, which is in fact linked with the BAS (e.g., Harmon-Jones & Sigelman, 2001). Harmon-Jones (2003) reported strong associations between measures of BAS and both psychometric and behavioral measures of aggression (also see Carver, 2004; and Wingrove & Bond, 1998, for similar BAS-aggression effects). According to RST, aggressive behavior is appetitively motivated as it provides a means to approach a state of relief or satisfaction. Somewhat counter-intuitively, this holds even if the response is prompted by a negatively valenced stimulus, in which case the behavior would be described as BAS-mediated active-avoidance (Gray, 1987b).

Tests of the Model

Gray's postulations, regarding the BIS and BAS in particular, gave rise to a number of specific hypotheses for personality research: Measures of Impulsivity should correspond to behavioral indexes of reward-reactivity, and measures of Anxiety should correspond to behavioral indexes of punishment-reactivity. "Reactivity" in this sense applies to some behavioral, affective, or motivational change elicited by reinforcing stimuli (see Pickering & Gray, 2001; or Smillie, Dalgleish, & Jackson, 2005). Thus, one would predict that impulsive individuals are motivationally affected by signals of reward (e.g., as indicated by the development of a preference for a rewarded choice option; Smillie & Jackson, 2006), show positive emotional reactions to signals of reward (e.g., as indicated by self-reported happiness following some indication of reward; Carver & White, 1994), and show superior learning under a reward schedule (e.g., as indicated by increases in performance on a feedback- and reward-guided category-learning task; Pickering, 2004b). Complementary predictions should hold for anxiety in the case of punishment signals. The previous account

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

is also prescriptive for the design and validation of RST questionnaires. First, such measures should capture susceptibility to reward and vulnerability to punishment.[1] Second, the putative physical independence of the BIS and BAS and the orthogonal trait representation depicted in Figure 1 have been taken to suggest that measures of Anxiety and Impulsivity should be unrelated (e.g., Carver & White, 1994; Torrubia et al., 2001). Third, as Impulsivity is located between Extraversion and Neuroticism, and Anxiety between Introversion and Neuroticism, this alignment should be reflected in concurrent validation investigations. In particular, BAS should be composed of 2 parts Extraversion to 1 part Neuroticism; BIS should be 2 parts Neuroticism to 1 part Extraversion (see Pickering et al., 1999).

As it turns out, all of these predictions have met with a perplexing variety of empirical findings, suggesting that some aspects of the model may require rethinking (Corr, 2001, 2004; Jackson, 2003; Matthews & Gilliland, 1999; Pickering et al., 1997). These include, but are not restricted to; the location of impulsivity and anxiety in factor space (Jackson, 2003; Pickering et al., 1999; Rusting & Larsen, 1999), the independence of the BIS/BAS systems and their associated personality traits and emotional states (e.g., Carver, 2004; Ball & Zuckerman, 1990; Pickering, 1997; Zinbarg & Revelle, 1989), the role of cognitive intermediaries, such as expectancies (Corr, 2002b; Zinbarg & Revelle, 1989), and the extent to which the trait model suggested by Gray corresponds to neuropsychological systems comprising the BIS and BAS (especially in the case of Impulsivity; Depue & Collins, 1999; Pickering, 2004b; Pickering & Gray, 2001; Smillie, 2005). While these various issues were surfacing in human personality research, however, a major revision of the structures and processes fundamental to RST was underway, based upon new empirical data in animal research (Gray & McNaughton, 1996, 2000). As these changes were separate to the problems identified during tests of RST as an explanation of personality, they seemed only to add to the growing uncertainty.

Refinement of the BIS, BAS, and FFS Functions

Gray and McNaughton's (2000) revised RST emerged as part of ongoing evaluation of both old and new animal-based data concerning the neuropsychology of anxiety, and the revised model is held to account for more of this data than any competing theory (McNaughton & Gray, 2002). To date, despite efforts to elucidate and draw attention to these revisions (e.g., Corr, 2002a, 2004; McNaughton & Corr, 2004), there has been minimal consideration of the potential implications that these theoretical refinements may have for personality measurement within the RST framework. Indeed, many prominent researchers in this area continue to conceptualize RST only in terms of the earlier version of the model (e.g., Ávila & Torrubia, 2004; Claridge & Davis, 2003; Gomez & Gomez, 2002).

In Gray and McNaughtons's (2000) revised model, the BAS still functions as a reward system (or positive feedback loop), but, in contrast with the original version of RST, it is now thought to mediate responses to all appetitive stimuli (not just conditioned). Similarly, the FFS is thought to mediate responses to all aversive stimuli (not just unconditioned), and, as such, adopts the punishment system role previously bestowed upon the BIS. The FFS has also been renamed the FFFS, denoting a Fight Flight and Freezing System (Gray & McNaughton, 1996, 2000). Here, theory development has greatly drawn upon animal research by D. C. Blanchard and Blanchard (1990a, 1990b), who report consistent displays of fight, flight, or freezing in response to threat. Flight and freezing are unlearned responses to distal threat stimuli, and fight to proximal threat stimuli. Specifically, a threatened animal will "fight" if the source (e.g., a predator) is within the immediate visio-spatial field, or has actually made physical contact. As distance from the predator increases the intensity of the response decreases; flight results if escape is physically possible, freezing if it is not.

As per the old RST, the new model focuses chiefly upon the BIS, which is still understood to be the neuropsychological epicentre of anxiety. However, while previously serving to regulate responses to conditioned aversive stimuli, it is now activated by sources of conflict; any inputs that might activate both the BAS and the FFFS. As such, the BIS is no longer a punishment system, but rather a conflict detection and resolution device. BIS-engagement inhibits ongoing behavior (both BAS and FFFS mediated), while simultaneously directing arousal and attentional resources toward the source of the conflicting stimuli. In the Gray-McNaughton (2000) revision, the resulting behavioral repertoire is referred to as "defensive approach," and is characterized by risk assessment and caution, refining Gray's earlier concept of passive avoidance. The BIS resolves conflict by engaging either the BAS (to continue approaching) or FFFS (to escape) according to some evaluative function of the reinforcing signals received. According to Gray and McNaughton (2000), the conflict resolution process always favors FFFS engagement by increasing the per-

[1] As opposed to typical experience of punishment (Fowles, 1987). Logically, an anxious person who is—according to RST—motivated to avoid punishment (BIS+) should tend not to bring themselves in contact with punishment to the same extent, or with the same frequency, as their less anxious counterparts (BIS–). Therefore, although individuals should have heightened responsivity to punishment, they are unlikely to experience these responses very often.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

ceived negative valence of inputs to this system (i.e., increasing the perceived intensity of threat). As such, a subject tends to be relieved of the experience of anxiety through some form of escape behavior.

Clearly, Gray and McNaughton's (2000) revisions have had minimal impact upon the way the BAS is conceptualized by personality psychologists. In fact, because the unconditioned/ conditioned distinction has been largely ignored in the design of various BAS scales (e.g., Jackson & Smillie, 2004; Torrubia et al., 2001), the psychometric and neuropsychological views of this system might now be in closer agreement. Furthermore, although changes to the prescribed functioning of the BIS appear particularly striking, they are not completely contrary to previous understanding of RST (Corr, 2002a; McNaughton & Corr, 2004). BIS-mediated reactions to punishment—that is, the interruption of behavior (and the increase of arousal, information gathering, and information processing, as described by Gray, 1982b)—have always been understood, if only implicitly, to require the presence of some other conflicting stimulus (as noted by Fowles, 1987, p. 419). In the case of classic extinction behavior, a rat eventually stops pressing a lever for food if the reward is terminated. Here the animal's frustration is not created by the absence of a reward per se, but by the mismatch between the "expectation" of reward and the experience of nonreward. Similarly, in the case of classic passive-avoidance, a rat inhibits ongoing behavior (e.g., running to a goal box) in response to a signal of threat, which conflicts with the motivation of the original behavior. One cannot escape the logic that, for a stimulus to elicit behavioral inhibition, the behavior in question must be motivated or actually occurring in the first place. Gray and McNaughton (2000) discuss various kinds of conflict, including not only those of appetitive and aversive stimulus pairings, but also those involving two mutually incompatible aversive stimuli (where one stimulus is less punishing and therefore relieving), and even conflicts between rewards (where one stimulus is less rewarding and therefore frustrative).

The major implication of Gray and McNaughton's (2000) revisions, so far as trait psychology is concerned, relates to the distinction between "fear" and "anxiety." Although anxiety was always said to be a result of BIS activation, and fear the result of FFFS activation (e.g., Gray, 1982b, 1987b), in practice the BIS had implicitly subsumed both responsibilities. Now the role of the FFFS has been given greater attention as an emotional system: Gray and McNaughton (2000, pp. 43–44) stated that threats (or potential threats) that need not be approached, need only produce escape or defense behavior (flight/fight/freeze). If approach is involved then BIS-mediated inhibition results due to conflict/mismatch. Otherwise, it would seem that FFFS and BAS alone are responsible for the active regulation of behavior (with continual passive monitoring by the BIS).[2] FFFS-mediated freezing is clearly distinguished from behavioral inhibition as it is an unlearned, fixed-action pattern, accompanied by a distinct posture (observed in animal subjects), and is resistant to blocking of behavior inhibition processes by the administration of anxiolytics or lesioning of the septo-hippocampal area (e.g., Gray & McNaughton, 1983; McNaughton & Gray, 1983). Conversely, the administration of panicolytics or lesioning of the amygdala will block FFFS manifestations (Gray & McNaughton, 1996; Hamm & Weike, 2005).

In terms of emotion, the distinction between "freezing" and "behavioral inhibition" parallels the differentiation of "fear" and "anxiety" (R. J. Blanchard, Griebel, Henire, & Blanchard, 1997). Unfortunately, although pharmacological and direct lesion methods can separate these emotions it is difficult to do so at other levels of analysis. For instance, in clinical settings the symptoms of panic disorder (putatively relating to the FFFS) and anxiety disorder (putatively relating to the BIS) show similar overlap. This is because higher levels of anxiety can trigger panic and the experience of panic can condition anxiety (Gray & McNaughton, 2000, pp. 327–331). In experimental psychophyisology, there is ambiguity in identifying an autonomic response (e.g., potentiated startle) as reflecting FFFS-mediated fear or BIS-mediated anxiety (Gray & McNaughton, 2000, pp. 44–45). The difficulties in distinguishing between fear and anxiety are also reflected by the fact that they have been treated as equivalent constructs in the wider psychology literature (White & Depue, 1999).

For these reasons, one might expect psychometric measures of fear and anxiety to be inextricably overlapped. In his critique of the Gray-Wilson Personality Questionnaire (GWPQ; Wilson, Barrett, & Gray, 1989), one psychometric measure of RST, Jackson (2003) demonstrated that flight, whose items were constructed to represent "pure fear" situations, is as good a predictor of trait anxiety as are the two BIS components (i.e., the passive avoidance and extinction scales). Similar conclusions were drawn by the original authors of the GWPQ, when principal components analysis evidenced factorial splitting and complexity among putative BIS and FFS items (Wilson, Gray, & Barrett, 1990). White and Depue (1999), however, cited considerable evidence from outside the immediate RST-based personality literature that supports a psychometric fear/anxiety distinction. A large part of this data uses Tellegen's

[2]Note that even if the BAS and FFFS are activated, a significant imbalance in the intensity of rewarding and punishing inputs will not create conflict. Rather, the BAS or FFFS will secure control over behavior. BIS-mediated anxiety is, therefore, perhaps most typical in environments or situations within which there is uncertainty or ambiguity regarding potential reward or punishment.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

(1982) Multidimensional Personality Questionnaire, which includes a broad-focus scale of Negative Emotionality (NEM; similar to Eysenck's "Neuroticism" and Gray's "Anxiety") and a narrow-focus scale of Harm Avoidance (HA; similar to Gray's notion of "Fear"). These two scales form separate factors and have a near-zero bivariate relationship. White and Depue (1999) also reported a clear divergence between these two measures when predicting pupillary reactivity to stimuli that they argue relate, respectively, to anxiety and fear: NEM predicts pupil dilation in response to darkness, whereas HA predicts pupil dilation in response to a Norepinephrine alpha-1 receptor agonist.

A likely reason for the discrepancy between results based on the GWPQ compared with the MPQ concerns item content. Specifically, existing measures of BIS-mediated anxiety would, if they include items referring to pure punishment situations (previously related to the BIS but now to the FFFS), also reflect aspects of fear. Therefore, RST questionnaires that are based upon the earlier perspective, such as the GWPQ, may include scales that confound the BIS and the FFFS. This problem might not apply to Tellegen's measures—indeed, Gray and McNaughton's (2000) revisions may have brought the RST view closer to that of Tellegen in this respect: The HA scale concerns reactions to unconditioned and conditioned physical punishers (which seems highly consistent with the undertakings of the FFFS), whereas the NEM scale reflects sensitivity to uncertainty, social comparison, and failure of one's efforts (which seems somewhat similar to the concept of "conflict," albeit couched within a social context).[3] White and Depue's (1999) thorough review and promising experimental data may therefore provide preliminary support for, and guidelines for further examination of, the fear/anxiety distinction emphasized in the new RST.

Overall, it seems reasonable to conclude that a revision of existing BIS and FFFS measurement inventories is in order. Presently, so-called BIS scales may either reflect a mixture of fear and anxiety, or may fail to wholly capture the notion of conflict. The development or identification of scales whose items clearly reflect either conflict and mismatch detection situations (BIS), or pure punishment and fearful threat situations (FFFS), is therefore essential for updating the RST trait model in line with Gray and McNaughton's (2000) revisions. A potential means to achieve this has been explored by Perkins and Corr (2006), through the extension of a novel self-report-based paradigm for assessing animal equivalents of fear and anxiety in human subjects (D. C. Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001). D. C. Blanchard and colleagues presented participants with vignettes describing various threatening situations, and asked them to choose from a set of responses analogous to rodent behavior. These scenarios varied in terms of dangerousness, escapability, proximity of threat, ambiguity of threat, and opportunity for protection. For example, "You are walking alone in an isolated but familiar area when a menacing stranger suddenly jumps out of the bushes to attack you," is a scenario with high danger and escapability, along with low ambiguity, close proximity, and opportunity for protection. Possible responses to these scenarios included putative FFFS output (e.g., "attack or struggle," "run away"), and BIS output (e.g., "check out, approach, or investigate"). Using this paradigm, D. C. Blanchard et al. (2001) showed that, in general, participants selected FFFS-related responses to pure threat situations and BIS-related responses to more ambiguous or only partly threatening scenarios.

Perkins and Corr (2006) sought to replicate this research, and then to map responses to these scenarios onto existing personality measures. All responses were ordinally coded in terms of defensive intensity and defensive direction. For instance, the response "run away" would be of intermediate defensive intensity oriented away from the threat. According to the distinctions RST now makes between FFFS-mediated fear and BIS-mediated anxiety, one would expect psychometric measures of fear to predict defensive avoidance (e.g., escape) whereas measures of anxiety should predict defensive approach (e.g., approach with caution, assess risk). Possible relationships between these measures and defensive intensity were not explicitly hypothesized, but examined for exploratory purposes. Results from a multiple regression showed that the Spielberger trait anxiety scale (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) predicted orientation toward the threat, $\beta = -.206, p < .05$, and lower defensive intensity, $\beta = -.202, p < .05$. Conversely, the Fear Survey Schedule (Wolpe & Lang, 1977) weakly predicted orientation away from the threat, $\beta = .146$, *ns,* and was not related to defensive intensity, $\beta = .028$, *ns*. A similar pattern was observed for Carver and White's (1994) BIS scale (defensive direction, $\beta = .276, p < .001$; defensive intensity, $\beta = .200$, *ns),* which was originally designed to measure the punishment system, and therefore, as we argue may be the case for many BIS measures in circulation, is likely to reflect the FFFS in the revised RST. These findings are generally consistent with the Gray-McNaughton (2000) revision, and suggest that D. C. Blanchard et al.'s (2001) threat scenarios may offer a promising avenue for improved trait conceptualization and measurement in RST.

[3]A reasonable argument could be made for the relevance of social stimuli to BIS mediated anxiety, due to the conflicting reinforcement signals that characterize such situations. That is, it seems highly unlikely that social exchanges are purely rewarding or punishing; rather, there might be clear perceptions of both (such as the potential for rejection during courtship). Relevant to this speculation is Macdonald and Leary's (2005) innovative linking of Gray and McNaughton's (2000) ideas to the explanation of social pain and rejection (but see Corr, 2005, for comment).

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

Interdependence among RST systems and corresponding personality traits. The next development to which we turn concerns clarifications of the traditionally assumed independence of the RST systems. Perhaps the best example of this general assumption is the view that measures of Anxiety and Impulsivity must be orthogonal. This interpretation may derive from the putatively distinct physiology of these systems (e.g., Gray, 1987b). It may also be related to Gray's description of RST traits as an alternative rotation of Extraversion and Neuroticism (Gray, 1970, 1981), thus anchoring the reward and punishment systems to two orthogonal personality axes. In either case, such details need not necessarily suggest that the outputs of these systems be independent or strictly separable. In fact, Gray and Smith's (1969) animal discrimination learning model, which provided one of the earliest blueprints for the structures and processes inherent to RST, explicitly represented the reward and punishment systems as mechanisms that compete (via mutual inhibition) for control over behavior. The implication here is that the effect of a stimulus on behavior depends not only upon the strengths of the stimulus and the reactivity of the system that it activates, but also upon the strength of the competing system(s). In other words, outputs of the BAS depend not only upon the reactivity of this system and the intensity of rewarding stimuli, but also upon the inhibitory strength of the BIS (and now, also of the FFFS).

The Gray and Smith (1969) model was explored by Pickering (1997) using neural network modeling to explicate the complexity of RST. Significantly, Pickering demonstrated that many empirical tests of RST that have apparently failed to support the model (as reviewed by Pickering et al., 1997) in fact yielded reasonable outcomes. That is, it seems in many cases it has been experimental predictions rather than results that have been inconsistent with RST. More recently, Pickering (2004a) revised the Gray-Smith model to incorporate the Gray-McNaughton (2000) revisions to RST discussed in the previous section (see Figure 2). Like the Gray-Smith model, there are mutual inhibitory pathways between the reward system (BAS) and the punishment system (formerly the BIS, now the FFFS). Functionally, these pathways imply that an output from the BAS is accompanied by an inhibitory output to the FFFS and vice versa. The key extension of the Gray-Smith model lay in the representation of the conflict-resolution system (i.e., the BIS), which is engaged only by the combined activation of the BAS and FFFS. As Gray and McNaughton specify that conflict-resolution is biased in favor of the FFFS, the BIS inhibits the BAS and (indirectly) excites the FFFS. The overall implication of the model is that behavior will, in mixed-incentive environments (i.e., most environments), reflect the combined influence of the three RST systems.

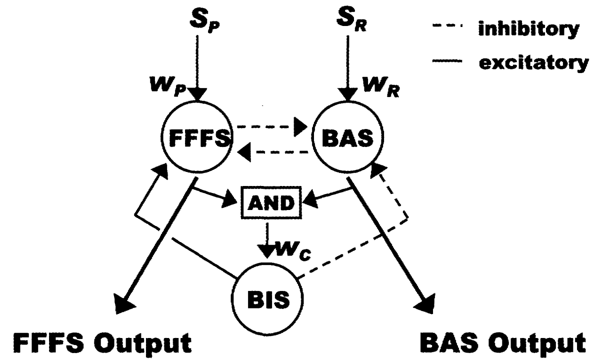

Figure 2. A dynamically interacting network model of new reinforcement sensitivity theory involving 3 systems. FFFS = Flight-Fight-Freeze (punishment) System. BIS = Behavioral Inhibition (conflict resolution) System. BAS = Behavioral Activation (reward) system. S_R and S_P are system inputs; w_R, w_P, and w_C are system reactivities. The AND representation of BAS/FFFS interaction in the activation of the BIS was instantiated as a multiplication, but is represented as AND to reflect that fact that both the BAS and FFFS need to be engaged to activate the BIS. It is important to note that the BIS only indirectly excites the FFFS, and thus the direct excitatory link shown in the model is for simplification only. Also, strictly speaking, it is the outputs of the FFFS and BAS that are directly influenced by the BIS, rather than the systems themselves (Gray & McNaughton, 2000, pp. 85–86), but for our purposes this distinction is not critical. For further details refer to the Appendix.

In an effort to rethink the set of predictions that constitute basic tests of RST, Corr (2001, 2002a) distinguished between the conventional focus upon separable effects of Gray's systems, and the joint effects implied by the Gray-Smith (1969) model and explored theoretically by Pickering (1997). The separable effects hypothesis predicts that responses to reward are independent of FFFS-reactivity and responses to punishment are independent of BAS-reactivity. Corr argued that separable effects should be observed when only reward or only punishment is delivered. They may also arise for individuals with hyper-reactive punishment or reward systems, or in mixed incentive conditions when either reward or punishment is of a high intensity (c.f. footnote 2). However, most individuals are presumably not BAS or FFFS hyper-reactive, and most environments comprise a mixture of stimulus valences (including controlled laboratory environments). In such cases, joint-effects of the reward and punishment systems are likely to occur. That is, effects of reward will partly depend upon FFFS-reactivity and effects of punishment will partly depend upon BAS-reactivity.[4]

[4]Corr (2002a) focused only upon the reward and punishment systems of RST, but we can logically extend his arguments to suggest that joint effects occur among all three systems as suggested by Pickering's (2004a) formal model (depicted in Figure 2). However, exactly how an independent or separable effect of the BIS could ever arise remains unclear to us, as this system is only activated by BAS-FFFS coactivation.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

Indeed, Gray (1973) noted that lesions that reduce responses to punishment also enhance responses to reward, and that direct simulation produces analogous effects (p. 425). In terms of personality, this means that reactions to reward, for instance, should depend not only upon Impulsivity, but also upon Anxiety and Fearfulness. Furthermore, due to the prevalence of conditions under which they may be expected to emerge, joint effects of RST systems may have greater overall explanatory power than either of their separate influences alone (as is the focus of conventional tests of RST). Statistically, these may emerge as main (additive) effects or interaction (multiplicative) effects. There is much evidence in the personality literature that seems to support such an implication, in terms of main effects (e.g., punishment system measures predicting reward system criteria) or interactions (e.g., BAS × BIS/FFFS effects on criteria). Exemplary instances of such effects have been observed by Ball and Zuckerman (1990), Corr (2002a), and Zinbarg and Revelle (1989). At the time they were observed, these findings were thought to be inconsistent with RST, but would now be seen as perfectly acceptable, a view that represents a major departure from earlier understanding of RST by most personality psychologists.

The clear implication for personality that Corr (2002a) derived from joint-effects in RST is that, for most individuals in most situations, personality traits will have codependent effects in tests or applications of RST. Less obvious is the implication of these system interdependencies for the mapping of Gray's motivational systems onto personality traits themselves. For instance; the typical interpretation of RST (i.e., separable effects) suggests that reactivity of the BAS (w_R in Figure 2) corresponds directly to a person's behavioral response to reward, which is in turn summarized by measures of trait Impulsivity. Figure 2 clearly shows, however, that behavioral output of the BAS is not simply a function of its system weight, but is rather determined by the combined activity of all three motivational systems. This raises the question of whether trait Impulsivity, assumed to reflect variation in BAS functioning, should be seen as relating to reactivity of the BAS or to the functional outcomes of the BAS (the latter of which is influenced by the activity of all three RST systems).

As we argue elsewhere (e.g., Pickering, 2004a), it seems very unlikely that people can directly introspect about their reinforcement reactivity weights. These weights function as amplifiers for neural signals elicited by reinforcing stimuli, and the suggestion that individuals consciously refer to the values of these weights while completing a personality questionnaire is biologically and cognitively implausible. It seems more likely, however, that a participant's responses to a personality questionnaire are based upon the functional outcomes of those weights. For example, items from Carver and White's (1994) BAS scale can easily be seen to reflect a person's assessment of the intensity of their reward-controlled behaviors (e.g., "When I see an opportunity for something I want, I get excited right away;" "When I want something I usually go all-out to get it"). Functional outcomes potentially indexed by such questionnaire items may include the mean output of the BAS, or the probability of the BAS gaining control of behavioral output in a rewarding context. Under the conventional separable subsystems view of RST, it would be safe to assume a strong relationship between functional outcomes and system reactivities, and reasonable, therefore, to suggest that measures of Impulsivity (indirectly) index the reactivity of the BAS (w_R). This is not the case with the understanding that the three RST systems have a combined effect upon behavior.

Assume then that psychometrically measured traits have some relationship to the functional outcomes of the RST systems (e.g., Impulsivity ≈ subjective rating of mean BAS output over a variety of situations). As this system outcome is itself nonindependent of activity in the other systems (due to excitation by the BIS and inhibition by the FFFS), then it may be possible for their corresponding personality traits also to be nonindependent. Such a thought may cause some irritation to those who have carefully developed uncorrelated RST scales, but, on the other hand, it would speak to the numerous instances of overlap among such measures in the literature (e.g., Carver & White, 1994; Smillie & Jackson, 2005; Wilson et al., 1989). For instance, the Appetitive Motivation scale (Jackson & Smillie, 2004) is a recent BAS measure whose validity does not appear to be compromised by a modest (but reliable) association with measures that could reflect functioning of Gray's punishment system ($r \sim -.20$).

Formal modeling by Pickering (2004a) lent some support to the validity of oblique RST traits. Trait data were simulated using a model of the form depicted in Figure 2 (technical details of the simulation appear in the Appendix). In these simulations, it was assumed that self-reported trait values (e.g., Impulsivity) would correspond to a functional outcome of one system only (e.g., BAS). In the example discussed here, the functional outcome used was the mean level of output (across varied situations) from each system. Mean BAS output might correspond, in subjective terms, to the mean level of an associated emotional state over a wide variety of circumstances. For each simulated subject, the model was allowed to run for 100 iterations, to ensure system activations would reach asymptotic values. The asymptotic values, however, were usually achieved in far fewer than 100 iterations. Data were then generated for 100 simulated subjects, with BIS/BAS/FFFS reactivity values (i.e., w_C, w_R, w_P) drawn randomly from independent normal distributions (for

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

each distribution, $M = 0$, $SD = 0.15$). The asymptotic system outputs generated for each subject were averaged across 200 stimulus events, each presenting some combination of reward and punishment signal (S_R and S_P varied across 200 simulated situations for each simulated subject; the values were drawn from independent uniform random distributions, with values in the range of 0–1). These situations would thus activate the BAS, FFFS and—when these were in direct conflict—the BIS, to widely varying degrees. Further technical details can be found in the Appendix.

The simulated trait data from this model demonstrated a strong negative correlation between "Impulsivity" and "Fear," $r = -.53, p < .001$, which is in agreement with the many instances in the literature of a negative relationship between measures of BAS and BIS (the latter of which, as we argued earlier, are likely to largely reflect FFFS reactivity). There was a strong positive relationship between "Anxiety" and "Fear," $r = .40, p < .001$, paralleling the surface overlap of these emotions in numerous contexts (described earlier). There was also a positive relationship between "Anxiety" and "Impulsivity," $r = .33, p < .001$, which seems counterintuitive, but is explained by the fact that BAS output is required for BIS activation (because the BIS is engaged by conflict between appetitive and aversive stimuli). Pickering (2004a) reported a series of regression analyses in which the functional outcomes from the model (mean system outputs) were predicted by the three system reactivities. Mean BAS output was independently influenced by w_P, w_R, and w_C, with the influence of w_R being positive, whereas w_P and w_C exerted negative influences, $R^2 = 0.89$. Mean FFFS output was well predicted by w_P alone (a positive association) with nonsignificant independent contributions from w_R and w_C, $R^2 = 0.82$. Mean BIS output was independently influenced by w_P, w_R, and w_C, with all three influences being positive, $R^2 = 0.85$.

These statistics emphasize that functional outcomes (such as the mean activation of the BAS over a range of situations) are not generally selectively dependent upon a single system. It is our contention that a person's responses on a personality questionnaire (as well as many other forms of measurable behavior) are likely to relate to functional outcomes of these systems and their interactions, rather than having a one-to-one correspondence with the system input sensitivities, as has previously been assumed or inferred. Thus, in our view, responses on a BAS-related personality trait measure are unlikely to be determined exclusively by w_R (the input reactivity of the BAS). If this assumption is correct, then the results of the modeling reported earlier suggests that the BAS trait measure may well be jointly and independently influenced by all three system reactivities: w_R (positively), along with w_P and w_C (both negatively). Follow-up simulations, allowing for alternative functional outcomes to provide the basis for trait dimensions (e.g., estimates of the probability of the BAS gaining control of behavior) produce substantively similar results (Pickering, in press). We feel this is a radical restatement of the way in which the key causal variables in RST (w_P, w_R, and w_C) map onto the traits concerned. Overall, these findings suggest that (a) questionnaires that are designed to have orthogonal measures of RST traits may require revision, and (b) questionnaires that contain oblique measures of RST traits may not be as problematic as the earlier view of RST might have appeared to suggest.

Reward-Reactivity Versus Trait Impulsivity: What is the BAS-Related Trait?

Although we have yet to do so in detail, it is worth briefly considering what implication the interplay of RST systems discussed in the previous section might have for the identification of RST traits themselves. For example, the modeling previously summarized suggests that high scores on a BAS-related trait may be typified by a combination of the following system sensitivities; high BAS, low FFFS, and low BIS. In his original formulation of the RST trait model, Gray (1973, 1981) maintained that Eysenck's trait of Extraversion reflects the relative balance of sensitivity in the reward and punishment systems. Indeed, as one moves along the Extraversion axis in Figure 1, reward system sensitivity increases while punishment system activity decreases. In contrast, when moving along the impulsivity diagonal reward system sensitivity increases strongly but punishment system sensitivity remains constant. Assuming that the true characterization of Extraversion and Impulsivity (in terms of reward and punishment system sensitivity) are as shown in Figure 1, this suggests that the trait hypothetically labeled Extraversion, rather than Impulsivity, may offer the best correspondence with functional outcomes of the BAS.

Although our suggestion regarding the correspondence of RST traits to functional outcomes is (presently) controversial, our speculation that the BAS-related trait may be identified as Extraversion, rather than Impulsivity, is in agreement with recent empirical data. Despite the fact that the BIS and FFFS were the less clearly defined systems in the original RST (as discussed earlier), tests of the BAS in personality research have actually produced the more discouraging findings (for reviews see Pickering & Gray, 2001; Pickering & Smillie, in press; Smillie, 2005). The key problem here concerns the ambiguity surrounding the BAS-related trait. Although the literature is replete with explicit assurances that "the trait reflecting the BAS is Impulsivity" (e.g., Gomez & Gomez, 2002, p. 1302), it is important to note that Gray linked the BAS to trait impulsivity on a relatively ad hoc basis

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

(Diaz & Pickering, 1993; Gray, personal communication). As Gray, Owen, Davis, and Tsaltas stated (1983), RST was

> first proposed (Gray, 1970) as a theory of anxiety … It was natural to suggest that, if anxiety reflects sensitivity to signals of punishment and nonreward, [an orthogonal rotation of this dimension] might correspond to sensitivity to signals of reward and non-punishment [and] might underlie the trait of high impulsivity. (pp. 184–185)

The clearly speculative nature of the BAS in terms of its correspondence with trait impulsivity can not be overemphasized.

A compelling argument that Impulsivity might have been a misnomer for the BAS-related trait was first provided by Depue and Collins (1999) in their neurobiological model of incentive motivation. They concurred with Gray's basic proposition that such motivational processes are a central cause of personality, but disagreed that appetitive motivation manifests as trait Impulsivity. Depue and Collins proposed that dopaminergic processes, widely regarded as central to BAS functioning (e.g., Derryberry & Tucker, 1991; Reuter, Schmitz, Corr, & Hennig, 2005), and commonly linked with trait Impulsivity or Psychoticism (Coscina, 1997; Gray, 1991; N. S. Gray, Pickering, & Gray, 1994; Pickering & Gray, 1999), are in fact most directly related to Extraversion. Among the evidence they present is a strong and significant relationship between dopamine receptor effects and Extraversion, but only modest relationships of these effects with Impulsivity (Depue & Collins, 1999, p. 513).[5] Admittedly, as Pickering (1999) observed, the overall balance of evidence is not tipped decisively in favor of Extraversion nor Psychoticism/Impulsivity as the primary correlate of dopaminergic neurotransmission. Nevertheless, this ambiguity might provide a reasonable explanation for why trait Impulsivity (aligned with Psychoticism), if not the BAS-related trait (aligned with Extraversion), might be mistakenly identified as such. It is possible that both Impulsivity/Psychoticism and Extraversion relate to distinct aspects of dopaminergic neurotransmission, but only Extraversion relates to the components that are more intimately associated with BAS functioning (as we speculate elsewhere; Smillie, Jackson, & Dalgleish, 2006).

Recent experimental evidence supports the view espoused by Depue and Collins (1999) regarding the identity of the BAS-related trait. Pickering (2004b) described a series of category-learning experiments in which a reliable association emerged between Psychoticism (measured using the Eysenck Personality Questionnaire, EPQ; Eysenck & Eysenck, 1975) and learning, but not when learning was contingent upon reward. Rather, it was EPQ Extraversion that seemed to best predict reward-contingent learning (and also reward-based motivational effects), suggesting that the BAS is more strongly related to this trait than to Impulsivity.[6] One particular finding from this review is especially striking; a "paired associate" version of a typical category-learning task was designed whereby each trial consisted of a stimulus presentation together with the category to which it belonged. Participants could therefore associate stimuli with categories without exposure to any explicit rewarding stimuli likely to engage the BAS (e.g., positive feedback). Learning was then assessed by a test phase, in which all stimuli were presented a second time and participants were required to assign to the correct category themselves. In this version of the task, EPQ-Psychoticism was significantly associated with learning, $r = 0.30$, $p = .02$, but EPQ-Extraversion was not, $r = 0.13$, $p = .21$. Results were then compared for a "BAS-engaging" version of the task. Here, participants learned by trial-and-error, with rewards (positive feedback and 10 UK pence) being given for each correct categorization. In the test phase after this version of the task, EPQ-Psychoticism was not associated with learning, $r = -0.001$, *ns*, but EPQ-Extraversion was, $r = 0.34$, $p = 0.015$. Overall, these results strongly suggest that the BAS-related trait is Extraversion.

Complementary evidence has been obtained more recently by Smillie and colleagues, who compared purpose-built BAS measures with generic trait Impulsivity scales in the prediction of reward-based learning and motivational effects (Smillie, 2005; Smillie et al., 2005; Smillie & Jackson, 2006). Smillie and Jackson (in press) found that three potential BAS measures, all strongly associated with Extraversion (BAS scale by Carver & White, 1994; Appetitive Motivation Scale by Jackson & Smillie, 2004; Functional Impulsivity scale by Dickman, 1990), predicted the development of a response-bias favoring the rewarded choice option in a go/no-go discrimination learning paradigm. A fourth measure, Dsyfunctional Impulsivity (defined by the scale's author as typical trait Impulsivity; Dickman, 1990, and highly correlated with Psychoticism, e.g., Chico, Tous, Lorenzo-Seva, & Vigil-Colet, 2003), did not converge with the other putative measures of BAS. Two further experiments using a category-learning

[5]As a caveat, we note that a key principle of Depue and Collins's (1999) argument is the perspective that Impulsivity emerges from the interaction of several biological systems, whereas Extraversion corresponds to a single substrate (i.e., the BAS, p. 497). Therefore, they may not necessarily agree with our view that, although the BAS-related trait corresponds to the output of the BAS only, this output itself is influenced by other biological systems.

[6]It is worth noting here that some research suggests that EPQ-Extraversion—unlike Eysenck's earlier measures of this trait—is almost entirely unrelated to trait Impulsivity (Rocklin & Revelle, 1981).

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

paradigm replicated and extended this research (Smillie et al., 2005). First, Carver and White's (1994) BAS-total scale predicted response bias in the same manner as observed in Smillie and Jackson's go/no-go discrimination task, but three measures of Impulsivity (these were the Impulsiveness, Risk-Taking, and Sensation-Seeking subscales of Psychoticism from the Eysenck Personality Profiler; Eysenck, Barrett, Wilson, & Jackson, 1992) did not. Second, the Carver and White BAS scale predicted learning (in terms of increases in response-sensitivity) when rewarding rather than punishing feedback was given, whereas a trait-Impulsivity scale (S. B. G. Eysenck, Pearson, Easting, & Allsopp, 1985; again, highly correlated with Psychoticism) did not. Overall, these experiments provide consistent evidence that measures of Impulsivity do not predict BAS-related learning and motivation criteria.

There has also been a recent burst of psychometric research to support the experimental work of Depue, Pickering, Smillie, and their colleagues. This work has distinguished Extraversion and purpose-built measures of BAS from Psychoticism and generic measures of Impulsivity. For example, Zelenski and Larsen (1999) factor analyzed psychometric measures relating to the theories of Gray, Eysenck, and Cloninger, and obtained three independent factors. The first comprised Neuroticism along with a number of measures relating to the BIS or anxiety. The second was composed of Extraversion along with a number of measures developed to assess the BAS (from both Cloninger's and Gray's theories). The third was composed of Psychoticism along with various measures associated with Impulsivity. Very similar results to these have since been obtained using factor analysis, structural equation modeling, and multivariate regression (Caseras, Ávlia, & Torrubia, 2003; Knyazev, Slobodskaya, & Wilson, 2004; Quilty & Oakman, 2004; Smillie & Jackson, in press; Smillie et al., in press). Finally, in some recent reviews of issues relating to clinical psychology, Dawe and colleagues argued that a distinction can be made between BAS-mediated behavior (which they call "reward-drive") and Impulsive behavior (which they call "rash-impulsiveness"), in connection with symptoms of eating disorders and substance abuse (Dawe, Gullo, & Loxton, 2004; Dawe & Loxton, 2004). Like the experimental and psychometric research reviewed here, the measures cited in these reviews as relating to "reward-drive" tend to be associated with Extraversion, whereas those related to "rash-impulsiveness" tend to be associated with Psychoticism.

The implication of all of this research for psychometric measurement in RST is clear: Extraversion, rather than Impulsivity, may reflect variation in BAS functioning. Considering the arbitrary nature with which Impulsivity was first identified as the BAS-related trait, and the success with which Extraversion has been employed as a proxy for the BAS prior to the availability of dedicated RST questionnaires (e.g., S. Gupta, 1990), such a conclusion is not so striking in hindsight. The potential theoretical implications of this view may require careful consideration, given that many processes associated with Extraversion are not entirely consistent with BAS functioning (e.g., differences in cortical arousal supposed to underlie Extraversion; Eysenck, 1967). On the other hand, a number of features of Extraversion are more adequately explained by drawing upon the notion of reward-reactivity. For instance, although Extraversion has a reliable association with positive affect (see Eysenck & Eysenck, 1985), no aspect of Eysenck's theory predicts this relationship (Corr, 2004). If Extraversion had a basis in reactivity of the BAS, however, the prediction of a relationship with positive affect would flow directly from Gray's postulates.

In any case, replacing Impulsivity with Extraversion (or some other trait) does not create any fundamental difficulty for RST that we can see. As we have argued throughout this article, the specific traits to which RST has been related are of no particular importance for the model—especially in the case of Impulsivity. Gray was not in any special way concerned with trait Impulsivity per se, but more so with traits that corresponded to dispositional appetitive motivation. (Equally, the only reason trait Anxiety is of importance to RST is because it describes the spectrum of dispositional physiology and psychological disorders associated with functioning of the BIS.) Although Gray (e.g., 1973, 1981) took great pains to distinguish Extraversion from the BAS-related trait, in terms of which dimension might best capture the lines of causal influence, one could argue that this distinction was entirely hypothetical. That is to say, the important feature of the BAS-related trait as an alternative to Eysenck's dimension was not its descriptive proximity to Extraversion—whether this is a 45° rotation (Gray, 1981), a 30° rotation (Pickering et al., 1999), or no rotation at all. As emphasized by Corr (2004, pp. 319–320), the important point was that a motivational explanation of personality overcame certain difficulties with the arousal-drive perspective advocated by Eysenck. If it so happens that Eysenck was correct about the fundamental role of Extraversion in the description of personality, but possibly mistaken about its underlying causes, then the bottom line of RST remains.

Conclusions

In his criticisms of the atheoretical approach adopted by advocates of the "Big Five" personality taxonomy, Hans Eysenck (1992) once quoted Imman-

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

uel Kant as having said that "measurement without theory is blind … just as theory without measurement is lame" (p. 672). Whatever reaction the Big-Fivers may have to the first sentiment, the latter is a criticism that could well be leveled at the advocates of RST. Measurement of the traits corresponding to the biological hardware and neurological processes described in Gray's model has been identified as a major agenda for RST research in personality (e.g., Pickering et al., 1997). In this article, we have argued that a challenge facing psychometric measurement of Gray's model relates to the fact that RST is not a theory of specific traits, but a broad framework of motivation, emotion, and learning. Therefore, basic research concerning the structure and organization of the biological systems of RST should inform psychometric measurement and trait conceptualization more strongly than might usually be the case for a personality model. Here we have considered three particular refinements, these being (a) Gray and McNaughton's (2000) revision to the basic details of RST, (b) formal considerations of dynamic interactions among the RST systems, and (c) the growing indication that the BAS-related trait may be represented by Extraversion rather than Impulsivity. These revisions have been driven by the accumulation of new data and clarification of Gray's original postulates, and, it is encouraging to note, often appear to bring RST closer to alternative perspectives in personality research. Table 1 summarizes the "old" and the "new" RST in terms of the issues we have considered in this article.

From this review and theoretical analysis, we draw a number of implications for psychometric measurement in RST. First, we identify the primary psychometric implication of Gray and McNaughton's (2000) "New RST" as of relevance to the distinction between BIS-mediated Anxiety and FFFS-mediated Fearfulness. We suggest that many measures of Anxiety may reflect a mixture of BIS and FFFS behavior, and consequently provide confounded measurement. We therefore conclude that existing measures of BIS/Anxiety and FFFS/Fearfulness require substantial revision, otherwise new measures will have to be devised. Fortunately, some guidelines for how such a revision might take place are provided by Tellegen's (1982) Multidimentional Personality Questionnaire, which seems capable of distinguishing between Fear and Anxiety. Recent work by Perkins and Corr (2006), based upon an animal ethoexperimental paradigm, is also promising. Next, in our consideration of system interdependence in RST (extending Corr, 2001; and Pickering, 1997, 2004b), we suggested that functional interdependence of the RST systems may result in nonindependence of traits corresponding to RST. A reevaluation of the importance of orthogonal trait measurement is therefore required, and it may be the case that oblique RST scales provide the most plausible trait representation of RST. Finally, in our review of recent research questioning the relationship of trait Impulsivity to the BAS (or reward-reactivity), we concurred with recent arguments (e.g., Depue & Collins, 1999) that variations in Gray's reward system correspond to measures of Extraversion. Full and proper evaluation of this possibility may not be trivial, given that the colossal literature on Extraversion may be difficult to fully reconcile with a "BAS ≈ Extraversion" hypothesis. At the same time, however, a number of features of Extraversion are more successfully explained in terms of reward-motivation than in terms of arousal (Corr, 2004).

As should always be the case in good science, our conclusions and the arguments upon which they are based must withstand close examination; as such they should be seen as a set of hypotheses to now be tested. We particularly emphasize the speculative and radical nature of our suggestion that RST traits correspond to functional outcomes, rather than system reactivities, of the BIS, BAS, and FFFS. A valuable step in evaluating

Table 1. *Summary of Revisions to Reinforcement Sensitivity Theory (RST).*

	Old RST	New RST
Behavioral Inhibition System		
Stimulus Input	CS–	Conflict (e.g., CS+ *with* CS–)
Behavioral Output	Avoidance	Defensive Approach
Trait Manifestation	Anxiety	Anxiety
Behavioral Activation System (or Behavioral Approach System)		
Stimulus Input	CS+	UCS+, CS+
Behavioral Output	Approach	Approach
Trait Manifestation	Impulsivity	Extraversion (???)
Fight, Flight, and Freezing System		
Stimulus Input	UCS–	UCS–, CS–
Behavioral Output	Avoidance	Defensive Avoidance
Trait Manifestation	???	Fearfulness
Trait Relationships	Uncorrelated	Correlated

Note: CS+ = conditioned signals of reward or relieving nonpunishment; CS– = conditioned signals of punishment or frustrative nonreward; UCS+ = unconditioned reward or relieving nonpunishment; UCS– = unconditioned punishment or frustrative nonreward; ??? = unknown.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

our perspective on this matter would be experimental differentiation of system reactivities and functional outcomes. For instance, Corr (2002a) suggested that RST systems will exert separable effects in pure reward or punishment scenarios (or when reward or punishment is of high intensity, or when subjects have heightened reactivity on one system). Therefore, one might correlate a BAS-related trait measure with a BAS-mediated behavioral response, elicited under pure reward versus mixed-incentive conditions. If the trait measure is more strongly associated with the behavior in the mixed-incentive condition, a case might be made that the measure reflects functional outcomes of the BAS. Alternatively, if the trait measure is more strongly associated with the behavior in the pure reward condition, one might conclude that the measure reflects reactivity of the BAS. It may be the case that some measures correspond to functional outcomes, whereas others index the latent system reactivities (against the view we have proposed). Knowing what properties of the underlying systems are reflected by the various potential RST measures seems certain to have implications for their use in future RST research.

As a final caveat, we note that this review has required the synthesis of a great variety of research that is often not in happy agreement. It is possible, therefore, that our attempts to summarize various theoretical perspectives and empirical findings suffer in places from lack of clarity, accuracy, or internal consistency. This in itself would be diagnostic of the complexity of RST, and would underline our view that personality research in particular suffers from a simplistic interpretation of Gray's model. The purpose of this review was to inform the development of better trait conceptualization and psychometric measurement in RST research, and we have hopefully provided some potential avenues for doing so. However, it seems that much more work may be needed to clearly define the constructs and behaviors of interest before we, as a research community, can be confident in the validity of such measures.

References

Ávila, C., & Torrubia, R. (2004). Personality, expectations, and response strategies in multiple-choice question examinations in university students: A test of Gray's hypotheses. *European Journal of Personality, 18,* 45–59.

Ball, S. A., & Zuckerman, M. (1990). Sensation seeking, Eysenck's personality dimensions and reinforcement sensitivity in concept formation. *Personality and Individual Differences, 11,* 343–353.

Blanchard, D. C., & Blanchard, R. J. (1990a). An ethoexperimental analysis of defence, fear, and anxiety. In N. McNaughton & G. Andrews (Eds.), *Anxiety* (pp. 188–99). Dunedin, New Zealand: Otago University Press.

Blanchard, D. C., & Blanchard, R. J (1990b). Anti-predator defence as models of animal fear and anxiety. In P. F. Brain, S. Parmigiani, R. J. Blanchard, & D. Blanchard (Eds.), *Fear and defence* (pp. 89–105). Chur, Switzerland: Harwood Academic.

Blanchard, D. C., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals. *Neuroscience & Biobehavioral Reviews, 25,* 761–770

Blanchard, R. J., Griebel, G., Henrie, J. A., & Blanchard, D. C. (1997). Differentiation of anxiolytic and panicolytic drugs by effects on rat and mouse defense test batteries. *Neuroscience and Biobehavioural Reviews, 21,* 783–789.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Briggs, I., & Myers, P. B. (1980). *Gifts differing.* Palo Alto, CA: Consulting Psychologists Press.

Carver, C. S. (2004). Negative affects deriving from the behavioural approach system. *Emotion, 4,* 3–22.

Carver, C. S., & White, T. L. (1994). Behavioural Inhibition, Behavioural Activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333.

Caseras, X., Ávila, C., & Torrubia, R. (2003). The measurement of individual differences in Behavioural Inhibition and Behavioural Activation Systems: A comparison of personality scales. *Personality and Individual Differences, 34,* 999–1013.

Cattell, R. B. (1950). *Personality; a systematic theoretical and factual study.* New York: McGraw-Hill.

Chico, E., Tous, J. M., Lorenzo-Seva, U., & Vigil-Colet, A. (2003). Spanish adaptation of Dickman's Impulsivity Inventory: Its relationship to Eysenck's Personality Questionnaire. *Personality and Individual Differences, 35,* 1883–1892.

Claridge, G., & Davis, C. (2003). *Personality and psychological disorders.* London: Arnold.

Cloninger, C. R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry, 44,* 573–590.

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,* 33–352.

Corr, P. J. (2002a). J. A. Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences, 33,* 511–532.

Corr, P. J. (2002b). J. A. Gray's reinforcement sensitivity theory and frustrative nonreward: A theoretical note on expectancies in reactions to rewarding stimuli. *Personality and Individual Differences, 32,* 1247–1253.

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioral Reviews, 28,* 317–332.

Corr, P. J. (2005). Social exclusion and the hierarchical defence system: Comment on Macdonald and Leary (2005). *Psychological Bulletin, 131,* 231–236.

Coscina, D. V. (1997). The biopsychology of Impulsivity: Focus on brain serotonin. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment and treatment* (pp. 95–115). London: Guilford Press.

Costa, P. T., & McCrae, R. R. (1992). Reply to Eysenck. *Personality and Individual Differences, 13,* 861–865.

Dawe, S., Gullo, M., & Loxton, N.J. (2004) Reward drive and rash impulsiveness as dimensions of impulsivity: Implications for substance misuse. *Addictive Behaviours, 29,* 1389–1406.

Dawe, S., & Loxton, N. J. (2004). The role of impulsivity in the development of substance use and eating disorders. *Neuroscience and Biobehavioural Reviews, 28,* 343–351.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioural and Brain Sciences, 22,* 491–569.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

Derryberry, D., & Tucker, D. M. (1991). The adaptive base of the neural hierarchy: Elementary motivational controls on network function. In R. Dienstbier (Ed.), *Nebraska Symposium on Motivation: Vol. 38. Perspectives on motivation* (pp. 289–342). Lincoln: University of Nebraska Press.

Diaz, A., & Pickering, A. D. (1993). The relationship between Gray's and Eysenck's personality spaces. *Personality and Individual Differences, 15,* 297–305.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology, 58,* 95–102.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, IL: Charles C. Thomas.

Eysenck, H. J. (1992). Five reasons four factors are not basic. *Personality and Individual Differences, 13,* 667–673.

Eysenck, H. J., Barrett, P., Wilson, G. D., & Jackson, C. J. (1992). Primary trait measurement of the 21 components of the P-E-N System. *The European Journal of Psychological Assessment, 8,* 109–117.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum Press.

Eysenck, H. J., & Eysenck, S. B. G. (1968). The measurement of Psychoticism: A study of factor stability and reliability. *British Journal of Social and Clinical Psychology, 7,* 286–294.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *The Eysenck Personality Questionnaire.* London: Hodder & Stoughton.

Eysenck, H. J., & Gudjonsson, G. H. (1989). *The causes and cures of criminality.* New York: Plenum Press.

Eysenck, S. B. G., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences, 6,* 613–619.

Fowles, D. C. (1987). Application of a behavioural theory of motivation to the concepts of anxiety and impulsivity. *Journal of Research in Personality, 21,* 417–435.

Fowles, D. C. (1993). Biological variables in psychopathology: A psychobiological perspective. In P. B. Sutker & H. E. Adams (Eds.), *Comprehensive handbook of psychopathology* (2nd ed., p. 5782). New York: Plenum Press.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicates and non-medicated parkinsonism. *Journal of Cognitive Nuerocience, 17,* 51–72.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. D. Butcher (Eds.), *Advances in personality assessment: Vol. 1* (pp. 203–234). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gomez, A., & Gomez, R. (2002). Personality traits of the behavioural approach and behavioural inhibition systems: Associations with processing of emotional stimuli. *Personality and Individual Differences, 32,* 1299–1316.

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

Gray, J. A. (1973). Causal models of personality and how to test them. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory* (pp. 409–463). London: Academic Press.

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A model for personality* (pp. 246–276). Berlin: Springer.

Gray, J. A. (1982a). Précis of the neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system. *Behavioural and Brain Sciences, 5,* 469–534.

Gray, J. A. (1982b). *The neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system.* New York: Oxford University Press.

Gray, J. A. (1987a). The neuropsychology of emotion and personality. In S. M. Stahl, S. D. Iverson, & E. C. Goodman (Eds.), *Cognitive neurochemistry* (pp. 171–190). Oxford, England: Oxford University Press.

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

Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement. Perspectives on individual differences* (pp. 105–128). New York: Plenum.

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

Gray, J. A., & McNaughton, N. (1983). Comparison between the behavioural effect of septal and hippocampal lesions: A review. *Neuroscience and Biobehavioral Reviews, 7,* 119– 88.

Gray, J. A., & McNaughton, N. (1996). The neuropsychology of anxiety: Reprise. In D. A. Hope (Ed.), *Nebraska Symposium on Motivation, 1995: Vol. 43. Perspectives on anxiety, panic, and fear: Current theory and research in motivation* (pp. 61–134). Lincoln: University of Nebraska Press.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety.* Oxford, England: Oxford University Press.

Gray, J. A., Owen, S., Davis, N., & Tsaltas, E. (1983). Psychological and physiological relations between anxiety and impulsivity. In M. Zuckerman (Ed.), *The biological bases of sensation seeking, impulsivity, and anxiety* (pp. 189–217). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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.

Gray, N. S., Pickering, A. D., & Gray, J. A. (1994). Psychoticism and dopamine D2 binding in the basal ganglia using single photon emission tomography. *Personality and Individual Differences, 17,* 431–434.

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

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology, 57,* 5–14.

Harmon-Jones, E. (2003). Anger and the behavioural approach system. *Personality and Individual Differences, 35,* 995–1005.

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: Evidence that insult-related relative left prefrontal activity is associated with experienced anger and aggression. *Journal of Personality and Social Psychology, 80,* 797–803.

Jackson, C. J. (2003). Gray's reinforcement sensitivity theory: A psychometric critique. *Personality and Individual Differences, 34,* 533–544.

Jackson, C. J., & Smillie, L. D. (2004). Appetitive motivation predicts the majority of personality and an ability measure: A comparison of BAS measures and a re-evaluation of the importance of RST. *Personality and Individual Differences, 36,* 1627–1636.

Knyazev, G. G., Slobodskaya, H. R., & Wilson, G. D. (2004). Comparison of the construct validity of the Gray-Wilson Personality Questionnaire and the BIS/BAS scales. *Personality and Individual Differences, 37,* 1565–1582.

Macdonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? *Psychological Bulletin, 131,* 202–223.

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.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

McAndrew, C. & Steel, T. (1991). Gray's Behavioural Inhibition System: A psychometric examination. *Personality and Individual Differences, 12,* 157–171.

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

McNaughton N., & Gray, J. A. (1983). Pavlovian counterconditioning is unchanged by chlordiazepoxide or by septal lesions. *Quarterly Journal of Experimental Psycchology, 35B,* 221–233.

McNaughton, N., & Gray, J. A. (2002). "The neuropsychology of anxiety" as it really is: A response to O'Mara (2001). *Neuropsychological Rehabilitation, 12,* 363–367.

Perkins, A., & Corr, P. J. (2006). Reactions to threat and personality: Psychometric differentiation of intensity and direction dimensions of human defensive behavior. *Behavioral and Brain Research, 169,* 21–28.

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

Pickering, A. D. (1999). Personality correlates of the dopaminergic facilitation of incentive motivation: Impulsive sensation seeking rather than extraversion? *Behavioural and Brain Sciences, 22,* 534–535.

Pickering, A. D. (2004a, July). *A formal model of "new" reinforcement sensitivity theory.* Paper presented at the 12th European Conference on Personality, Groningen, The Netherlands (for a copy of this talk, see http://homepages.gold.ac.uk/aphome/ecp12.ppt).

Pickering, A. D. (2004b). The neuropsychology of impulsive antisocial sensation seeking personality traits: From dopamine to hippocampal function? In R. M. Stelmack (Ed.), *On the psychobiology of personality: Essays in honour of Marvin Zuckerman.* London: Elsevier.

Pickering, A. D. (in press). Formal and computational models of reinforcement sensitivity theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality.* Cambridge, England; Cambridge University Press.

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). Amsterdam, The Netherlands: Pergamon/Elsevier.

Pickering, A. D., & Gray, J. A. (1999). The neuroscience of personality. In L. Pervin & O. John (Eds.), *Handbook of personality* (2nd ed., pp. 277–299). New York: Guilford Press.

Pickering, A. D., & Gray, J. A. (2001). Dopamine, appetitive reinforcement, and the neuropsychology of human learning: An individual differences approach. In A. Eliasz & A. Angleitner (Eds.), *Advances in research on temperament* (pp. 113–149). Lengerich, Germany: PABST Science Publishers.

Pickering, A. D., & Smillie, L. D. (in press). The behavioural activation system: challenges and opportunities. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality.* Cambridge, England: Cambridge University Press.

Quilty, L. C., & Oakman, J. M. (2004). The assessment of Behavioural Activation—The relationship between Impulsivity and Behaviour Activation. *Personality and Individual Differences, 37,* 429–442

Reuter, M., Schmitz, A., Corr, P., & Hennig, J. (2005). Molecular genetics support Gray's personality theory: The interaction of COMT and DRD2 polymorphisms predicts the behavioural approach system. *The International Journal of Neuropsychopharmacology, 1,* 1–12.

Revelle, W. (1995). Personality processes. *Annual Review of Psychology, 46,* 295–328.

Rocklin, T., & Revelle, W. (1981). The measurement of extraversion: A comparison of the Eysenck Personality Inventory and the Eysenck Personality Questionnaire. *British Journal of Social Psychology, 20,* 279–284.

Rusting, C. L., & Larsen, R. J. (1999). Clarifying Gray's theory of personality: A response to Pickering, Corr & Gray. *Personality and Individual Differences, 26,* 367–372.

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

Smillie, L. D. (2005). *Trait representation of J. A. Gray's Behavioural Activation System: Distinguishing between Reward-Reactivity and Impulsivity.* Unpublished doctoral thesis, University of Queensland, Australia.

Smillie, L. D., Dalgleish, L. I., & Jackson, C. J. (2006). Personality, learning and motivation: Two tests of J. A. Gray's reinforcement sensitivity theory. *Personality and Individual Differences, 40,* 1039–1050.

Smillie, L. D., & Jackson, C. J. (2005). The Appetitive Motivation scale and other BAS measures in the prediction of Approach and Active Avoidance. *Personality and Individual Differences, 38,* 981–994.

Smillie, L. D., & Jackson, C. J. (2006). Functional Impulsivity and Reward-Reactivity. *Journal of Personality, 74,* 47–83.

Smillie, L. D., Jackson, C. J., & Dalgleish, L. I. (in press). Conceptual distinctions among Carver and White's (1994) BAS scales: A Reward-Reactivity versus trait Impulsivity perspective. *Personality and Individual Differences.*

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *STAI Manual.* Mountain View, CA: Consulting Psychology Press.

Tellegen, A. (1982). *Brief manual for the Multidimensional Personality Questionnaire.* Unpublished manuscript, University of Minnesota, Minneapolis.

Torrubia, R., Ávila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's Anxiety and Impulsivity dimensions. *Personality and Individual Differences, 31,* 837–862.

White, T. L., & Depue, R. A. (1999). Differential association of traits of fear and anxiety with norepinephrine and dark induced pupil reactivity. *Journal of Personality and Social Psychology, 77,* 863–877.

Wilson, G. D., Barrett, P. T., & Gray, J. A. (1989). Human reactions to reward and punishment: A questionnaire examination of Gray's personality theory. *British Journal of Psychology, 80,* 509–515.

Wilson, G. D., Gray, J. A., & Barrett, P. T. (1990). A factor analysis of the Gray-Wilson Personality Questionnaire. *Personality and Individual Differences, 11,* 1037–1045.

Wingrove, J., & Bond, A. J. (1998). Angry reactions to failure on a cooperative computer game: The effect of trait hostility, behavioural inhibition, and behavioural activation. *Aggressive Behaviour, 24,* 27–36.

Wolpe, J., & Lang, P. J. (1977). *Manual for the Fear Survey Schedule.* San Diego, CA: Educational and Industrial Testing Service.

Zelenski, J. M., & Larsen R. J. (1999). Susceptibility to affect: A comparison of three personality taxonomies. *Journal of Personality, 67,* 761–791.

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

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012

Appendix

The model presented in Figure 2 and discussed in the text is specified as a set of differential equations that specify the change in activation of the three systems of RST (BAS, FFS, and BIS). These equations are numerically simulated via equivalent difference equations that are iterated across repeated timesteps. The key model equations are shown here:

$$dBAS/dt = (M_E - \max(BAS,0)) * S_R * w_R - (M_I + BAS) * (K_4 * FFFS^{out} + K_2 * BIS^{out}) - K_3 * BAS$$

$$dFFFS/dt = (M_E - \max(FFFS,0)) * S_P * w_P - (M_I + FFFS) * \max((K_4 * BAS^{out} - K_2 * BIS^{out}),0) - K_3 * FFFS$$

$$dBIS/dt = (M_E - BIS) * FFFS^{out} * BAS^{out} * w_C - K_3 * BIS$$

where *FFFS*, *BAS*, and *BIS* are the activations of the systems concerned with $FFFS^{out}$, BAS^{out}, and BIS^{out} being the output of those systems. The sensitivities of these systems are w_P, w_R, and w_C respectively. S_P and S_R are the strengths of the input punishment and reward stimuli respectively. Other equation parameters are defined later.

The excitatory activation of the BAS is driven by the product of the reward stimulus strength (S_R) and the BAS sensitivity (w_R). An analogous product term is used to activate the FFFS. The effect of this input is iterated over the timesteps along with the other effects acting on system activations (see following). The activations of the BAS, FFFS, and BIS in the model are nonlinear. In particular, the activation of system values is bounded by a maximum excitation value (given by constant $M_E = 1.0$) and activations (of BAS and FFS) can be inhibited below zero (down to a maximally inhibited activation value of –0.5, determined by the constant $M_I = 0.5$). The outputs of the systems (e.g., BAS^{out}, which are just the activations of the systems thresholded at zero) act on other systems to determine their activation levels. Each system has a small passive decay constant ($K_3 = 0.05$) as well.

The equations reveal that the actions of the BIS output on the BAS and FFFS activations were implemented in an asymmetric way. The equations show that BIS^{out} sums with the inhibitory influence of $FFFS^{out}$ to inhibit BAS activation. However, BIS^{out} acts to inhibit the inhibitory effect of BAS^{out} and thus has a net (indirect) excitatory effect upon FFFS activation. This feature was adopted so as to somewhat reduce the influence of the BIS on FFFS activation (this is beneficial for reasons explained later). An alternative approach would have been to allow the BIS to excite the FFFS directly, by having BIS^{out} sum with the excitatory input to the FFFS ($= S_F * w_F$). We have explored this alternative method and the results were similar.

Another feature is that the influence of the BAS on the FFFS (and vice versa) is scaled by parameter K_4. The value chosen (= 0.2) was such that the BAS and FFFS were not especially good at resolving the competition between themselves, allowing the effects of BIS output to be visible across a range of contexts. With larger values of K_4 (e.g., 0.8) the more active system (of the BAS or FFFS) can almost always suppress the weaker system's activity and output toward or below zero. This obviates the need for the conflict resolving action of the BIS and also makes the BIS input (which is driven by $BAS^{out} * FFFS^{out}$) almost always close, or equal, to zero.

A further important feature is that the influence of the BIS on the activation of the other systems was scaled by parameter K_2 relative to the direct inputs to the system. The value of K_2 (= 0.2) was chosen to be small so that the effect of the BIS, which biases in favor of FFFS activation, is not excessive. Larger values of K_2 result in an undesirable situation where the input to the BAS ($= S_R * w_A$) can be considerably larger than that to the FFFS ($= S_F * w_F$), and yet the modest BIS activation that results can still tip the final values of the system activations so that FFFS > BAS. Finally, it should be noted that the AND representation of BAS/FFFS interaction in the activation of the BIS (see Figure 2) was instantiated as a multiplication operation. All values of the model parameters were chosen so that the system interactions performed "sensibly" and stably for certain individual subjects (i.e., for specific values of w_R, w_P, and w_C) under key stimulus conditions.

The simulation described in this article, along with other variations of this model, is shown in the slides from the presentation by Pickering (2004a), available on the web at http://homepages.gold.ac.uk/aphome/talks.html. The model itself is available as MATLAB source code at http://homepages.gold.ac.uk/aphome/simulations.html. A more complete exploration of this subject using formal models is to be published shortly (Pickering, in press).

Downloaded from psr.sagepub.com at UNIV OF VIRGINIA on October 6, 2012