Pers Soc Psychol Bull OnlineFirst, published on March 15, 2007 as doi:10.1177/0146167206296951

Distinguishing Between Learning and Motivation in Behavioral Tests of the Reinforcement Sensitivity Theory of Personality

Luke D. Smillie
Goldsmiths, University of London

Len I. Dalgleish
University of Stirling

Chris J. Jackson
University of Queensland

According to Gray's (1973) Reinforcement Sensitivity Theory (RST), a Behavioral Inhibition System (BIS) and a Behavioral Activation System (BAS) mediate effects of goal conflict and reward on behavior. BIS functioning has been linked with individual differences in trait anxiety and BAS functioning with individual differences in trait impulsivity. In this article, it is argued that behavioral outputs of the BIS and BAS can be distinguished in terms of learning and motivation processes and that these can be operationalized using the Signal Detection Theory measures of response-sensitivity and response-bias. In Experiment 1, two measures of BIS-reactivity predicted increased response-sensitivity under goal conflict, whereas one measure of BAS-reactivity predicted increased response-sensitivity under reward. In Experiment 2, two measures of BIS-reactivity predicted response-bias under goal conflict, whereas a measure of BAS-reactivity predicted motivation response-bias under reward. In both experiments, impulsivity measures did not predict criteria for BAS-reactivity as traditionally predicted by RST.

Keywords: *RST; learning; motivation; Signal Detection Theory; Gray*

Concepts from the neuroscience of learning and motivation are being increasingly recognized for their potential to provide a major explanatory framework for personality. Theories that explain trait variation in such terms, by positing biobehavioral systems that regulate motivation and mediate reinforcement learning, are all centrally linked with J. A. Gray's (1970, 1973, 1981, 1987) Reinforcement Sensitivity Theory (RST) of personality (Fowles, 2006). The seminal work on this model (e.g., Gray, 1973) was credited as a revolution in personality research (Depue & Collins, 1999), and today, RST is seen as providing a full-fledged neuropsychology of temperament (Corr, 2004; Matthews & Gilliland, 1999). Comprehensive evaluation of the model has, however, proved notoriously difficult (see Corr, 2001; Pickering et al., 1997) and sophisticated paradigms for evaluating RST have been highlighted as a key agenda for research in this area. In an attempt to meet this demand, the present article draws a distinction between learning and motivation

Authors' Note: Preparation of the article was facilitated by a grant from the Economic and Social Research Council (PTA-026-27-0708) awarded to Luke D. Smillie. We are indebted to Alan Pickering of Goldsmiths College, University of London, and to William Revelle of Northwestern University for their insightful comments and advice on this research. Reprint requests should be sent to Luke D. Smillie, Department of Psychology, Goldsmiths, University of London, New Cross, London, SE14 6NW; e-mail: l.smillie@gold.ac.uk.

PSPB, Vol. X No. X, Month XXXX xx-xx
DOI: 10.1177/0146167206296951

processes in behavioral predictions derived from RST. A Signal Detection paradigm is then developed to test those predictions throughout two experimental investigations.

RST: Background and Challenges

The central features of Gray's RST are a Behavioral Activation System (BAS) and a Behavioral Inhibition System (BIS). In its original form (Gray, 1973, 1982), the BAS was understood to mediate responses to conditioned signals of reward and the BIS to mediate responses to conditioned signals of punishment. In its recently revised form (Gray & McNaughton, 2000) the BAS is activated by *all* forms of reward and the BIS is activated by punishments "*when they must be approached*" (Gray & McNaughton, 2000, p. 84), a stimulus input referred to as goal conflict. According to both the newer and older versions of RST, the outputs of the BIS are (a) interruption of ongoing behavior, (b) negative affect, (c) direction of attention toward the source of conflict, and (d) increased arousal. These outputs facilitate resolution of goal conflict by assessing potential threat and, if necessary, avoiding behavior that would bring the subject into contact with the threat. Putative outputs of the BAS are (a) invigoration of ongoing behavior, (b) positive affect, (c) direction of attention toward the source of reward, and (d) increased arousal. These outputs act to maintain or strengthen rewarded behavior, thus bringing the subject into closer spatiotemporal contact with reward.

According to Gray's (1981) application of RST to the explanation of personality variation, the functioning of the BIS and BAS systems underlie two major trait dimensions. Specifically, BIS-reactivity has been linked with individual differences in trait anxiety, whereas BAS-reactivity has been linked with individual differences in trait impulsivity. Although this proposal has been widely embraced, more recent literature has challenged its validity, particularly regarding the BAS = impulsivity view (see Dawe & Loxton, 2004; Quilty & Oakman, 2004; Smillie & Jackson, 2006; Zelenski & Larsen, 1999). To avoid confusion, this article shall distinguish BIS-reactivity and BAS-reactivity from the traits to which they are presumed to relate. Central to the RST view of personality is the premise that a BIS-reactive person is more sensitive or reactive to goal conflict, relative to a BIS-nonreactive person. Conversely, a BAS-reactive person is more sensitive or reactive to rewarding stimuli than is a BAS-nonreactive person (Gray & McNaughton, 2000; Pickering & Gray, 2001).

It must be realized that the theoretical impact of RST greatly exceeds the empirical evidence in its favor to date. RST has fundamentally influenced the direction of personality psychology as a whole (Fowles, 2006), but tests of Gray's central postulates have produced a bewildering array of findings (see Corr, 2001, 2004; Matthews & Gilliland, 1999; Pickering, 2004; Pickering et al., 1997; Pickering & Gray, 2001). Unexpected findings have enabled refinement of or clarification to the model, a necessary process for any theory, and one that has possibly enriched rather than weakened RST. Nevertheless, most of the reviews cited here identify imprecise or ambiguous predictions (e.g., Pickering, 1997, in press) and problematic experimental tests (e.g., Corr, 2001) as the major contributors to the mixed findings obtained to date; as Corr (2001) notes, "in several, crucial respects, RST has yet to be adequately tested" (p. 348). A critical agenda for RST is therefore to derive careful predictions from, and sophisticated tests of, the fundamental tenets of RST.

Learning and Motivation Processes in Behavior: Two Predictions From RST

RST is first and foremost a behavioral theory, grounded in animal ethology. In this article, we wish to distinguish between two kinds of behavioral prediction inherent to the model. The first prediction relates to what might be called the learning component of RST: Increases in arousal and attention toward the target stimulus, as a result of BIS or BAS activation, are thought to increase information processing and thereby facilitate learning (Corr, Pickering, & Gray, 1997; Gray, 1987; Pickering & Gray, 2001). As a result, BIS-reactive individuals should display superior learning in situations of goal conflict (compared to those with a less reactive BIS), whereas BAS-reactive individuals should display superior learning when their behavior is rewarded (compared to those with a less reactive BAS). This learning component of RST can be distinguished from what might be called the motivational component of RST, giving rise to a second prediction: Behavioral inhibition resulting from BIS activation is thought to facilitate (passive) avoidance of potential threat. Conversely, behavioral invigoration resulting from BAS activation is thought to facilitate approach of the rewarding stimulus. These effects are understood to be motivational in nature because they concern the direction of behavior rather than some absolute performance output and arise due to the emotional consequences of a response for the subject (Gray & Smith, 1969; Pickering & Gray, 2001; Schultz, 1998). In personality terms, BIS-reactive individuals (compared to those with a less reactive BIS) should more readily avoid or be more strongly disposed against responses that bring the subject into goal conflict, such as approaching a stimulus associated with punishment. Conversely, BAS-reactive individuals (compared to those with a less reactive BAS) should more readily approach or be more strongly disposed toward responses that are rewarded.

In the broader behavioral neuroscience literature concerned with effects of reinforcement, the distinction often is made between learning and motivation processes (e.g., Schultz, 1998, p. 1). However, this has not been previously emphasized in RST (with the exception of Pickering & Gray, 2001, p. 115). This has potentially significant consequences for interpreting previous research because in certain experimental situations the learning component of RST may oppose the motivational component. For instance, suppose that in a go/no-go task, incorrect go responses (responses to distractor) are punished and the number of correct go responses (responses to target), relative to a baseline block, are examined as the criterion. This experiment should produce BIS activation because good performance requires making a response that is associated with punishment (i.e., approaching punishment) and thus the subject is exposed to goal conflict. From the learning component of RST, one might predict BIS-reactive individuals to improve performance more rapidly than their BIS-nonreactive counterparts and, therefore, make more correct go responses relative to baseline. However, from the motivational component of RST, one also could predict BIS-reactive individuals to avoid the go response due to its association with potential threat, resulting in fewer correct go responses (due to fewer go responses being made overall). Contradictory predictions arising from the two components of RST may therefore help to explain some of the mixed findings in the literature (as Pickering & Gray, 2001, p. 126, have demonstrated by reanalysis of previous research) and are prescriptive for future tests of RST.

A potentially powerful means to distinguish between the learning and motivation consequences of BIS- and BAS-mediated behavior is provided by Signal Detection Theory (SDT; Macmillan & Creelman, 1991; McNicol, 1972; Swets, Tanner, & Birdsall, 1961). For binary choice tasks, such as described in the go/no-go example above, SDT analysis provides two behavioral criteria. The first is response-sensitivity, or *d*-prime (d'), and measures the extent to which an individual can discriminate one objective state from another (e.g., target vs. distractor). Increases in d' indicate improvements in discrimination ability and, therefore, positive learning of the task. The second criterion is response bias, or beta (β), and measures an individual's disposition toward one response or another (e.g., go vs. no-go). Changes in response-bias indicate shifts in a person's motivation to respond in a particular direction (McNicol, 1972). Therefore, one might expect that predictions from the learning component of RST might be effectively assessed in terms of response-sensitivity, whereas predictions from the motivational component of RST might be effectively assessed in terms of response-bias.

Overview of Experiments

In this article, we provide two behavioral tests of RST using a Yes/No categorization task, and SDT analysis is employed to distinguish between learning and motivation processes. To evaluate the learning component of RST, Experiment 1 examines changes in response-sensitivity as a function of (psychometrically measured) BIS- and BAS-reactivity. It is predicted that BAS-reactivity will predict the development of greater response-sensitivity under rewarding conditions (Hypothesis 1), whereas BIS-reactivity will predict the development of greater response-sensitivity under conditions of goal conflict (Hypothesis 2). To evaluate the motivational component of RST, Experiment 2 examines changes in response-bias as a function of BIS- and BAS-reactivity. It is predicted that BAS-reactivity will predict the development of a response-bias indicating an increase in (or approach of) rewarded responses (Hypothesis 3), whereas BIS-reactivity will predict the development of a response-bias indicating a decrease in (or passive avoidance of) responses associated with goal conflict (Hypothesis 4). BIS- and BAS-reactivity will be assessed using a purpose-built RST questionnaire measure. In addition, to assess the generality of any relationships observed, generic measures of trait anxiety (putatively BIS-mediated) and trait impulsivity (putatively BAS-mediated) also will be administered.

EXPERIMENT 1: REINFORCEMENT SENSITIVITY AND LEARNING

Experiment 1 was designed to assess predictions from RST concerning learning (Hypotheses 1 and 2). Learning was defined here in terms of the rate of the increase of response-sensitivity throughout the experiment. The Yes/No categorization task employed for this study presents graphical summaries of the personal characteristics of hypothetical job candidates. Participants assumed the role of a personnel selection officer and made sequential Yes/No decisions (Select or Reject) as to whether they would select the candidate represented in each of the vignettes. This task has been shown in preliminary research to yield similar results to a classic signal detection paradigm (auditory detection) while providing a generic, relatively engaging, and archetypal decision-making situation (Smillie & Dalgleish, 2001). Key predictions are clarified in the Method section.

METHOD

Participants

Sixty students from the University of Queensland participated in this study. Of these, 50 were credited as part of a psychology course requirement and 10 were volunteers who had responded to advertisements (61% women, M age = 20.60, SD = 6.01).

Materials

Questionnaires

Participants completed the BIS/BAS scales (Carver & White, 1994). This consists of 13 items to assess reactivity of the BAS (CWBAS scale), which can be summed to form a total scale ($\alpha = .81$) or divided into three subscales known as Reward-Responsiveness, Drive, and Fun-Seeking ($\alpha = .73, .76, .66$, respectively). An additional 7 items provide measurement of BIS-reactivity (CWBIS scale; $\alpha = .74$). The BAS items concern behavioral and affective reactions to reward (e.g., "When I see an opportunity for something I like, I get excited right away"), whereas the BIS items concern behavioral and affective reactions to punishment (e.g., "I feel worried when I think I have done poorly at something"). All questions are answered on a 4-point Likert-type scale from 1 (*strongly agree*) to 4 (*strongly disagree*). Initial research using these measures (Carver & White, 1994) demonstrated that all three BAS scales predict self-reported happiness when anticipating reward, whereas the BIS scale predicts nervousness in anticipation of punishment.[1]

To measure trait impulsivity and anxiety, the anxiety scale of the State-Trait Anxiety Inventory–Form Y (STAI-Y; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Impulsiveness scale of the I^7 Impulsiveness Questionnaire (I^7; S.B.G. Eysenck, Pearson, Easting, & Allsopp, 1985) also were administered. The STAI-Anxiety scale consists of 20 items to which participants indicate on a 4-point scale (1 = *almost never*, 4 = *almost always*) how often they feel a particular way (e.g., "I feel nervous and restless"). The internal consistency of this scale is reported as $\alpha = .93$ (Spielberger et al., 1983) and there is a significant body of research to support its validity (e.g., Oei, Evans, & Crook, 1990). The I^7 Impulsiveness scale consists of 19 questions to which participants respond Yes or No (e.g., "Do you generally do and say things without thinking"). The internal consistency of this scale tends to be reasonable ($\alpha = .82 \sim .85$). Consistent with typical conceptualizations of impulsivity, the I^7 Impulsiveness scale is primarily correlated with Psychoticism (Eysenck et al., 1985).

Categorization Task

The categorization task presents graphical vignettes of hypothetical job candidates (e.g., see Figure 1), to which participants must respond serially. Each vignette provides ratings on five selection criteria (presentation/grooming, manner, communication ability, sales and service skills,

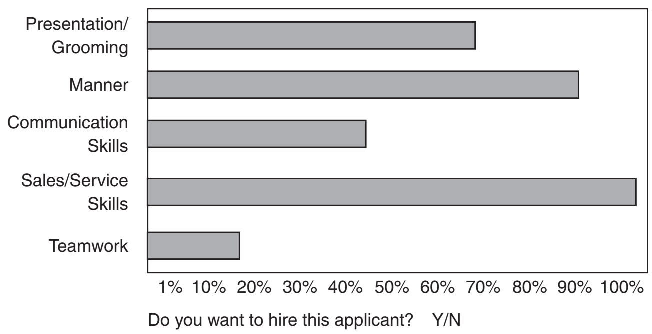

Figure 1 Sample vignette from category-learning task used in Experiments 1 and 2.

and teamwork skills). These ratings are depicted as separate bars on a single graph, each bar corresponding to an arbitrary scale from 0 (*completely unsuitable*) to 100 (*completely suitable*). Beneath the graphical display was the question, "Do you want to hire this applicant?" Participants responded to this question on a standard keyboard using the "Y" key for Yes and the "N" key for No. Each vignette was displayed for the length of time it took participants to respond, at which point a feedback message ("That was a correct decision" or "That was an incorrect decision") was displayed. Participants could then press any key to proceed to the next vignette.

The data on which the five numerical ratings depicted in the vignettes were based were generated as five normally distributed variables with a mean of 40 and a standard deviation of 18 (N = 125 cases). These provided the data for the nonsuitable applicants. Data for the suitable applicant profiles were then generated by incrementing the five variables by a constant of 15, thus creating an additional 125 cases with a mean of 55 and a standard deviation of 18. Accordingly, the two objective states of "suitable applicant" and "unsuitable applicant" were based on two overlapping normal distributions, one being a linear transposition of the other. This satisfies key assumptions of Signal Detection Theory regarding the nature of stimuli to be discriminated (Swets et al., 1961). Because the within-category stimulus variance was slightly larger than the between-category stimulus variance, the task was very difficult. To prevent implausible applicant vignettes (e.g., a score of 0 for communication skills but 100 for teamwork), correlations among the five cues were fixed at plausible values ($r \approx .30$) based on previous research (described by Smillie & Dalgleish, 2001).

Design

To assess RST predictions concerning learning, participants were randomly assigned to one of two experimental

TABLE 1: Descriptive Statistics and Intercorrelations for Psychometric Measures Used in Experiment 1

	M	SD	α	*Anx*	*Imp*	*CWBIS*
STAI-Anxiety	39.25	9.72	.91			
I^7 Impulsiveness	7.67	4.29	.75	–.04		
CWBIS	20.37	3.99	.79	.58**	–.19	
CWBAS	40.50	5.16	.77	–.45**	.32*	–.29*

NOTE: $N = 60$. Anx = Anxiety; Imp = Impulsiveness; CWBIS = Carver and White Behavioral Inhibition System; STAI = State-Trait Anxiety Inventory; CWBAS = Carver and White Behavioral Activation System. *$p < .05$. **$p < .01$.

conditions. Participants in the confirmatory feedback condition (coded as 0; $n = 30$) predominantly received feedback when they were correct, such that p(feedback|correct selection) = .80 and p(feedback|correct rejection) = .80, whereas p(feedback|incorrect selection) = .25 and p(feedback|incorrect rejection) = .25. Participants in the corrective feedback condition (coded as 1; $n = 30$) predominantly received feedback when they were incorrect, such that p(feedback|incorrect selection) = .80 and p(feedback|incorrect rejection) = .80, whereas p(feedback|correct selection) = .25 and p(feedback|correct rejection) = .25. Confirmatory feedback was expected to be rewarding, whereas corrective feedback was expected to introduce goal conflict because both response options will be associated with verbal punishment but must nevertheless be approached. All participants were instructed that they should use the feedback, when available, to guide their learning.

Measurement and Statistical Analysis

Formulae for calculating SDT parameters are widely available and comprehensively documented by Stanislaw and Toderov (1999). The formula for calculating response-sensitivity, where Z_{Sls} and Z_{Sln} are the standard normal values corresponding to the proportion of correct and incorrect Yes responses, is given as

$$d' = Z_{Sls} - Z_{Sln} \quad (1)$$

To assess learning in the two experimental groups, a prior-50 moving average of d' (d'_Δ) was calculated for each participant throughout the 250 experimental trials using a time-series analysis. That is, d' was computed for trials 1 to 50, then 2 to 51, and so on, for each participant in the experiment. This moving average was then correlated with trial number, the resultant coefficient giving an index of linear increases or decreases in performance throughout the experiment, $r_{d'\Delta.t}$.

To evaluate predictions from RST relating to learning, Moderated Multiple Regression (MMR) was employed (Cohen, Cohen, Aiken, & West, 2003). This is an appropriate analysis for evaluating interactions between continuous (e.g., trait) and dichotomous (e.g., experimental condition) variables. It enables a test of the prediction that learning will be positively predicted by measures of BAS-reactivity in the confirmatory feedback condition (Hypothesis 1) and by measures of BIS-reactivity in the corrective feedback condition (Hypothesis 2).

RESULTS

Preliminary Analyses

Table 1 shows means, standard deviations, internal consistency, and correlations among scales. CWBIS and STAI-Anxiety are positively correlated, whereas CWBAS and I^7 Impulsiveness are positively correlated (although this latter relationship is considerably weaker). This weaker correlation could be seen as consistent with claims that impulsivity is related to, but not synonymous with, BAS.

Preliminary inspection of the data showed that for most participants, d'_Δ decreased dramatically after (approximately) 200 trials. This could reflect fatigue or possibly a pattern of trials toward the end of the experiment that was experienced as being more difficult (possibly a more likely explanation given that experiments of these kind routinely employ 200-500 trials; Smillie & Dalgleish, 2001). Excluding the last 50 trials of the experiment from analyses resulted in a clear linear relationship between trials and d'_Δ for all participants. Thus, the dependent variable (DV) for this study was $r_{d'\Delta.t}$ throughout the first 200 trials of the experiment.

Manipulation Check

The theoretical rationale for the prediction that BIS will predict learning under corrective feedback, and BAS will predict learning under confirmatory feedback, derives from the putative increase in arousal and information processing resulting from BIS or BAS activation by affective stimuli. It is therefore necessary to show that confirmatory feedback was perceived as rewarding and that corrective feedback was perceived as punishing and therefore could be assumed to engage the BAS and BIS. Comparison of a postexperiment affect rating (How did the feedback you received about your decision making make you feel? 1 = *good*, 2 = *neutral*, 3 = *bad*) for participants in the two experimental conditions revealed that participants receiving corrective feedback reported feeling worse after feedback ($M = 2.37$, $SD = 0.81$), whereas those receiving confirmatory feedback reported feeling better ($M = 1.57$, $SD = .67$), $t(58) = 4.28$, $p < .001$.

Relationship of BIS and BAS Measures With Response-Sensitivity

To test key predictions relating RST to learning, two MMR analyses were conducted. All continuous variables including the criterion were standardized prior to analysis to reduce multicollinearity among predictors and to improve the interpretability of the solution. As such, the regression coefficients reported here are standardized and are thus indicative of effect size. The first MMR assessed the relationship that Carver and White's (1994) BIS/BAS scales had with learning in the two experimental conditions. Main effects were entered at Step 1 (CWBIS, CWBAS, condition; 0 = confirmatory, 1 = corrective) and interaction terms were entered at Step 2 (CWBIS × Condition, CWBAS × Condition). At Step 1, the model was statistically nonsignificant, $R^2 = .006$, $F < 1$, *ns*, and none of the predictors were significantly related to learning, all *t*s < 1. At Step 2 the model was significant, $R^2 = .19$, $F(5, 54) = 2.60$, $p = .04$, as was the increment in prediction, $R^2\ ch = .19$, $Fch(2, 54) = 6.32$, $p = .003$. The Condition × CWBAS interaction term reached significance,[2] B = .49, $t(54) = 2.06$, $p = .04$, as did the Condition × CWBIS interaction term, B = –.55, $t(54) = 2.29$, $p = .03$. Analysis of simple slopes (Jaccard, Turrisi, & Wan, 1990) indicated that CWBAS predicted increased response-sensitivity under confirmatory feedback, B = –.31, $t(54) = 2.06$, $p = .04$, but was not related to response-sensitivity under corrective feedback, B = –.18, $t(54) = 1.10$, $p = .28$. Conversely, CWBIS predicted increased response-sensitivity under corrective feedback, B = .36, $t(75) = 2.19$, $p = .03$, but was not related to response-sensitivity under confirmatory feedback, B = –.20, $t(54) = 1.06$, $p = .30$. These results are depicted in Figure 2.

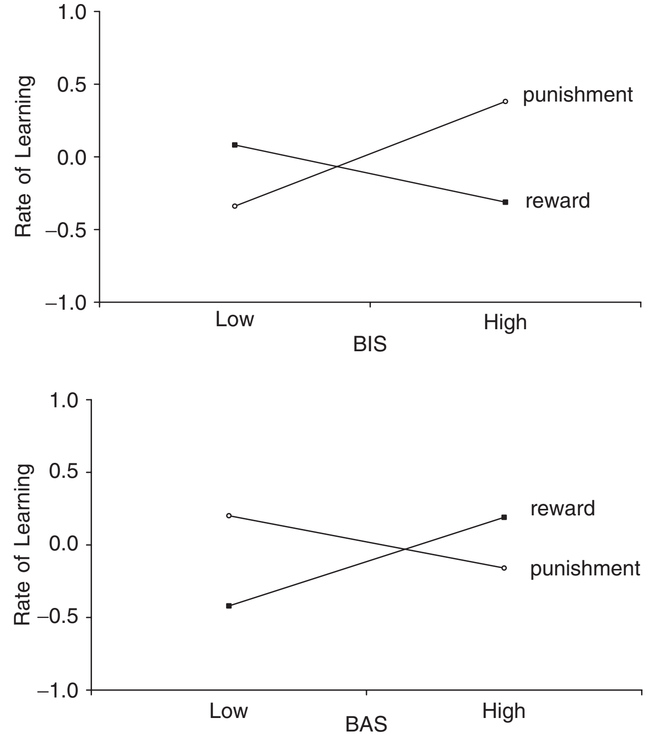

Figure 2 Plot of simple slopes demonstrating an interaction of (a) CWBIS and (b) CWBAS with condition in the prediction of learning (in terms of the rate of increase in response-sensitivity throughout trials).

NOTE: Rate of Learning ($r_{d'\Delta,t}$) is standardized and, as such, zero on the Y-axis corresponds to the mean rate of learning, which was .12 (reflecting the fact that this task was very difficult). Low and high values are defined as 1 *SD* below and above the mean, respectively. CWBIS = Carver and White Behavioral Inhibition System; CWBAS = Carver and White Behavioral Activation System.

The second MMR examined the generality of relationships observed for the BIS/BAS scales when using generic measures of trait impulsivity and anxiety to measure RST constructs. Again, main effects (IMP, ANX, condition) were entered at Step 1 and interaction terms (IMP × Condition, ANX × Condition) were entered at Step 2. At Step 1, the model was statistically nonsignificant, $R^2 = .09$, $F(3, 56) = 1.8$, $p = .15$; however, IMP was a significant (negative) predictor of learning, B = –28, $t(56) = 2.24$, $p = .03$. This indicated that higher scores on the I^7 Impulsiveness scale predicted a decrease in response-sensitivity throughout the experiment. At Step 2, the model still fell short of statistical significance, $R^2 = .16$, $F(5, 54) = 2.02$, $p = .09$, as did the increment in prediction, $R^2\ ch = .07$, $Fch(2, 54) = 2.17$, $p = .12$. Again, only IMP was a significant predictor of learning in the full model, again indicating that higher scores on this measure were negatively related to learning, B = –.44, $t(54) = 2.35$, $p = .02$.

To allow for the possibility of a relationship between ANX and learning that was not independent of the relationship between IMP and learning (or vice versa), two post hoc MMR analyses were conducted. The first (Step 1 = IMP, condition; Step 2 = IMP × Condition) verified that IMP was negatively related to learning, B = –.34, $t(56) = 1.78$, $p = .08$, but did not interact with condition in the manner predicted by Hypothesis 1, B = .34, $t(56) = 1.32$, $p = .19$. In the second MMR (Step 1 = ANX, condition; Step 2 = ANX × Condition), the interaction term for ANX × Condition approached significance, B = –.45, $t(56) = 1.66$, $p = .09$, and simple slopes analysis indicated that the relationship between ANX and learning approached significance under corrective feedback, B = .36, $t(56) = 1.61$, $p = .09$, and near-zero under confirmatory feedback, B = –.06, $t < 1$, *ns*. This finding suggests that the relationship that STAI-Anxiety may have with learning is similar to the CWBIS scale but considerably weaker and not independent of the relationship between I^7 Impulsiveness and learning.

Discussion

This experiment yields several important findings. First, Carver and White's (1994) BIS/BAS scales were related to learning in a Yes/No categorization task in a manner predicted directly by RST. BAS-reactivity as measured by the CWBAS scale predicted increased response-sensitivity throughout trials only when responses tended to be rewarded (Hypothesis 1), whereas BIS-reactivity as measured by the CWBIS scale predicted increased response-sensitivity throughout trials only when responses tended to be punished (Hypothesis 2). These results support the supposed link that BIS- and BAS-reactivity have with learning according to RST. Second, STAI-Anxiety did not interact with condition in the prediction of learning as would be expected of a measure of BIS-reactivity. Ancillary analyses suggested a weak relationship between STAI-Anxiety and learning under corrective feedback, but this relationship did not reach significance. This suggests that if STAI-Anxiety is related to BIS-reactivity in the same way as the CWBIS scale, then that relationship is a fragile one. Third, the only significant relationship involving I^7 Impulsiveness was a main effect, indicating that impulsivity was a negative predictor of learning. As such, this result does not support Gray's suggestion that the trait corresponding to BAS-reactivity might be identified as impulsivity and is broadly consistent with the view of impulsivity as having negative consequences for learning or performance (e.g., Barratt, 1985; Mathias & Stanford, 2003).

EXPERIMENT 2: REINFORCEMENT SENSITIVITY AND MOTIVATION

The second experiment in this article was designed to assess predictions from RST concerning motivation (Hypotheses 3 and 4). The task employed was identical to that used in the first experiment; however, feedback manipulations were designed to create an inequality in the motivation to respond "Yes" or "No." BAS-mediated approach and BIS-mediated avoidance were to be examined in terms of shifts in response-bias, which is intimately related to an individual's motivation to respond in a particular direction (McNicol, 1972). Predictions are clarified in the Method section after the experimental design is explained.

Method

Participants

Eighty-one undergraduate psychology students enrolled at the University of Queensland participated in this study in exchange for course credit (65% women; *M* age = 20.70, *SD* = 3.61).

Materials

Questionnaires

All participants completed the BIS/BAS scales (described in Experiment 1). Again, it was of interest to compare purpose-built RST scales with more general measures of impulsivity and anxiety, and as such, the Anxiety and Impulsiveness scales from the Eysenck Personality Profiler (EPP; H. J. Eysenck, Barrett, Wilson, & Jackson, 1992) were administered. EPP-Anxiety is a 20-item measure demonstrating reasonable internal consistency ($\alpha = .80 \sim .85$) and validity (Jackson, Furnham, Forde, & Cotter, 2000; Petrides, Jackson, Furnham, & Levine, 2003). High scorers tend to be worriers and are easily upset by negative events (e.g., "Are you inclined to get yourself all worked up over nothing?"). EPP-Impulsiveness ($\alpha = .75 \sim .80$) also consists of 20 items, many of which are from the I^7 Impulsiveness scale, and is defined in the EPP manual (H. J. Eysenck & Wilson, 2000) as the tendency to act spontaneously (e.g., "Do you often do things on the spur of the moment?"). All of scales from the EPP are responded to using a 3-point (*Yes/No/Can't Decide*) Likert-type scale.

Task Description

The category-learning task used to investigate predictions from RST concerning motivation was described in Experiment 1. However, although it was of interest to examine learning throughout the skill acquisition phase in Experiment 1, it was now of interest to assess the shift in response-bias from a neutral baseline. To provide baseline measurement of response-bias, during which participants could learn the task and develop a relatively neutral decision tendency, a further 250 trials were generated in the same manner described in Experiment 1. Therefore, a total of 500 trials were compiled for the two experimental blocks (baseline, treatment) in which all participants would serve.

Design

All participants completed the 250 baseline-block trials under full feedback; that is, each decision during the first experimental phase was followed by a text message on the screen that read either "That was a correct decision" or "That was an incorrect decision." It was expected that most participants would approach a near-neutral response-bias throughout the course of these full-feedback trials and thus be equally disposed to respond "Yes" or "No." For the treatment block, participants were randomly assigned to one of two experimental groups using a double-blind procedure. Half of the participants received predominantly confirmatory feedback for "Yes" responses (*N* = 41), whereas the other half received predominantly corrective feedback for "Yes" responses (N = 40). Specifically, for participants in the confirmatory feedback condition, p(feedback|correct selection) = .80 and p(feedback|incorrect selection) = .25, whereas for participants in the corrective feedback condition, p(feedback|correct selection) = .25 and p(feedback|incorrect selection) = .80. The probability for either group receiving confirmatory or corrective feedback for "No" decisions was fixed at .10. It was expected that these manipulations would elicit an increase in (or approach of) "Yes" response under confirmatory feedback and a decrease in (or passive avoidance of) "Yes" response under corrective feedback, as reflected by shifts in response-bias from block 1 to block 2.

Measurement and Statistical Analyses

The formula for calculating response-bias, where $Z_{S|s}$ and $Z_{S|n}$ are the standard normal values corresponding to the proportion of correct and incorrect "Yes" responses and e denotes an exponential function, is given as

$$\beta = {}_{e}\left\{\frac{Z^2_{S|n} - Z^2_{S|s}}{2}\right\} \quad (2)$$

where values greater than 1 indicate a response-bias toward responding "No" and values less than 1 indicate a response-bias toward responding "Yes."

In Experiment 1, rate of learning was examined by correlating a moving average of response-sensitivity with trial. This may not be suitable for observing changes in β, which unlike typical performance measures, does not linearly approach an asymptote. Rather, participants are able to adopt any initial response-bias and make increases or decreases toward an appropriate threshold. Furthermore, because observers typically have poorly calibrated response-bias (e.g., Kaernbach, 2001; Maddox & Bohil, 1998), trial-by-trial changes in β are likely to fluctuate around a mean value. An appropriate analysis to capture shifts in β would therefore calculate the mean level of β under a baseline block, which can then be compared with a block of trials in which the focal manipulation is made. It is then possible to calculate a standard error of β, SE_β, and thereby measure the standardized shift in β from baseline to test block, $Z_\Delta\beta$ (Smith, 1969). This measure would take into account variability around the mean of β in each experimental block.

The formulae for deriving SE_β and $Z_\Delta\beta$ are as follows: If n_s is the number of "Yes" cases and n_n is the number of "No" cases, whereas $P_{S|s}$ is the proportion of correct "Yes" responses and $P_{S|n}$ is the proportion of incorrect "Yes" responses, whereas $Z_{S|s}$ and $Z_{S|n}$ are the standard normal values corresponding to the proportion of correct and incorrect "Yes" responses and $O_{S|s}$ and $O_{S|n}$ are the ordinates of the standard normal distribution corresponding to the proportions of correct and incorrect "Yes" responses, then

$$SE_\beta = \beta^2\left[\frac{Z^2_{S|s}P_{S|s}(1-P_{S|s})}{n_s O^2_{S|s}} + \frac{Z^2_{S|n}P_{S|n}(1-P_{S|n})}{n_n O^2_{S|n}}\right] \quad (3)$$

and

$$Z_\Delta\beta = \frac{\beta_1 - \beta_2}{\sqrt{\dfrac{SE^2_{\beta_1} + SE^2_{\beta_2}}{2}}} \quad (4)$$

where negative values of $Z_\Delta\beta$ indicate a standardized shift in response-bias toward "Yes" and positive values indicate a standardized shift in response-bias away from "Yes" (Gourevitch & Galanter, 1967; Marascuilo, 1970). As for Experiment 1, focal analyses shall be conducted using Moderated Multiple Regression (MMR). These will enable a test of the prediction that shifts in the motivation to respond "Yes" will be predicted by measures of BAS-reactivity in the confirmatory feedback condition (Hypothesis 3) and by measures of BIS-reactivity in the corrective feedback condition (Hypothesis 4).

TABLE 2: Descriptive Statistics and Intercorrelations for Psychometric Measures Used in Experiment 2

	M	SD	α	*Anx*	*Imp*	*CWBIS*
EPP-Anxiety	16.35	8.47	0.82	—		
EPP-Impulsiveness	19.93	8.57	0.78	0.11	—	
CWBIS	19.77	4.05	0.80	0.56**	–0.20	
CWBAS	39.75	5.25	0.77	0.16	0.40**	.013

NOTE: N = 81. Anx = Anxiety; Imp = Impulsiveness; CWBIS = Carver and White Behavioral Inhibition System; EPP = Eysenck Personality Profiler; CWBAS = Carver and White Behavioral Activation System. **p < .01.

Results

Descriptive and Preliminary Statistics

Table 2 depicts means, standard deviations, internal consistency, and intercorrelations for the personality scales used in this experiment. These descriptive statistics are broadly similar to those observed in Experiment 1. As before, CWBIS is highly correlated with trait anxiety, whereas the correlation between CWBAS and trait impulsivity is slightly weaker.

Inspection of idiographic data confirmed that within-block shifts in β were not linear but converged on a mean value, from lower values of β for some subjects and from higher values for others. Similarly, scatterplots for a prior-50 moving average of beta throughout trials (within each experimental block) showed a clear failure of linearity. This confirms that a within-block analysis similar to that used in Experiment 1 would be inappropriate for examining changes in response-bias. As such, the dependent variable for Experiment 2 was the standardized shift in response-bias from the first block of 250 trials to the second block of 250 trials, as given by Equation 1. Unlike Experiment 1, and despite the longer duration of Experiment 2, potential fatigue effects were not observed.

Manipulation Check

The theoretical rationale for the prediction that measures of BIS-reactivity will predict development of a more stringent response-bias when "Yes" responses are associated with potential threat, and BAS-reactivity will predict development of a more liberal response bias when "Yes" responses are preferentially rewarded, derives from motivational effects of BIS- and BAS-reactivity elicited by goal conflict and reward. It is therefore necessary to show that the two experimental conditions were eliciting approach and passive avoidance.

A 2 (condition) × 2 (block) ANOVA, with β as the dependent variable, revealed no significant effect of block, $F < 1$, *ns*, a significant effect of condition, $F(1, 79) = 6.30$, $p = .014$, which was qualified by a significant Condition × Block interaction, $F(1, 79) = 9.06$, $p = .003$ (see Figure 3). Simple effects of condition at Blocks 1 and 2 indicated that there was no significant response-bias difference between the experimental groups in Block 1, $F(1, 79) = 2.28$, $p = .14$, but a significant difference in Block 2, $F(1, 79) = 8.88$, $p = .004$. Specifically, in the first experimental block, before participants were divided into groups, response-bias was almost exactly neutral, $\beta = .98$.[3] In the second experimental block, participants who received confirmatory feedback for "Yes" responses had developed a stronger preference for that response, $\beta = .83$, than participants who received corrective feedback for "Yes" responses, $\beta = 1.10$. These mean values are both on the expected sides of 1, which is the neutral value for β. Furthermore, both groups appear to have shifted their response-bias during the experiment, as indicated by the simple effects of block for each condition; that is, the confirmatory group made a significant negative shift in response-bias (indicating an increase in "Yes" responses), $F(1, 79) = 5.39$, $p = .023$, whereas the corrective group made a borderline-significant positive shift in response-bias (indicating a decrease in "Yes" responses), $F(1, 79) = 3.72$, $p = .05$. As such, it was concluded that the experimental manipulations were having their intended effect on the motivation to respond "Yes" in the categorization task.

Relationship of BIS and BAS Measures With Response-Bias

Two MMR analyses were conducted to test hypotheses relating to motivation, mirroring the procedure described in Experiment 1. The first MMR assessed the relationship of CWBIS and CWBAS with shift in response-bias, $Z_{\Delta}\beta$, under the two experimental conditions. As for Experiment 1, main effects were entered at Step 1 (CWBIS, CWBAS, and condition; 0 = confirmatory, 1 = corrective) and interaction terms were entered at Step 2 (CWBIS × Condition, CWBAS × Condition). At Step 1, the model was significant, $R^2 = .14$, $F(3, 77) = 4.13$, $p = .009$, with condition contributing significantly toward prediction, $B = .79$, $t(77) = 3.45$, $p = .001$, whereas CWBIS ($B = .15$) and CWBAS ($B = -.04$) did not contribute significantly, $p > .10$. Consistent with the manipulation check, the main effect of condition indicated that participants in the corrective feedback condition increased response-bias (i.e., made fewer "Yes" decisions, $M = .34$, $SD = .94$), whereas participants in the confirmatory feedback condition decreased response-bias (i.e., made more "Yes" decisions, $M = -.36$, $SD = 1.06$). At Step 2, the model was significant, $R^2 = .24$, $F(5, 75) = 4.25$, $p = .001$, as was the increment in prediction, $R^2\ ch = .10$, $Fch(2, 75) = 4.80$, $p = .011$. Again, condition contributed significantly toward prediction, $B = .78$, $t(75) = 3.52$, $p = .001$, whereas CWBIS ($B = -.066$) and CWBAS ($B = -.24$) did not, $p > .10$. In addition, the two interaction terms both contributed significantly, indicating that condition moderated the relationship of CWBAS, $B = .45$, $t(75) = 2.04$, $p = .04$, and CWBIS, $B = .47$, $t(75) = 2.09$, $p = .04$, with shift in response-bias.[4] Analysis of simple slopes (Jaccard et al., 1990) indicated that CWBAS predicted a shift in response-bias toward "Yes" when it was rewarded, $B = -.32$, $t(75) = 2.11$, $p = .03$, but was not related to response-bias when it was punished, $B = .14$, $t < 1$, *ns*. Conversely, CWBIS predicted a shift in response-bias away from "Yes" when it was punished, $B = .38$, $t(75) = 2.51$, $p = .01$, but was not related to response-bias when it was rewarded, $B = -.06$, $t < 1$, *ns*. These results are depicted in Figure 4.

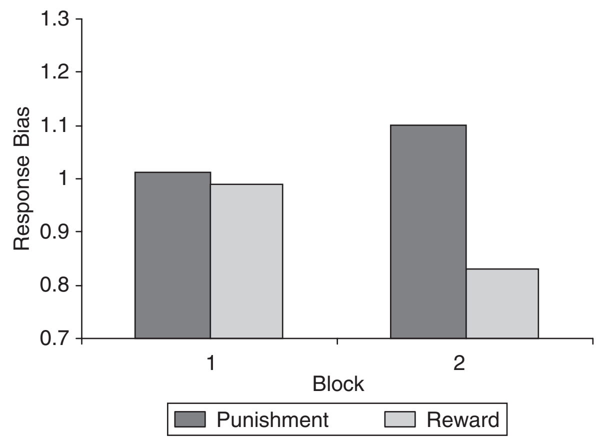

Figure 3 Results of a 2 × 2 ANOVA demonstrating that response-bias shifted to favor "Yes" in the confirmatory feedback group and "No" in the corrective feedback group.

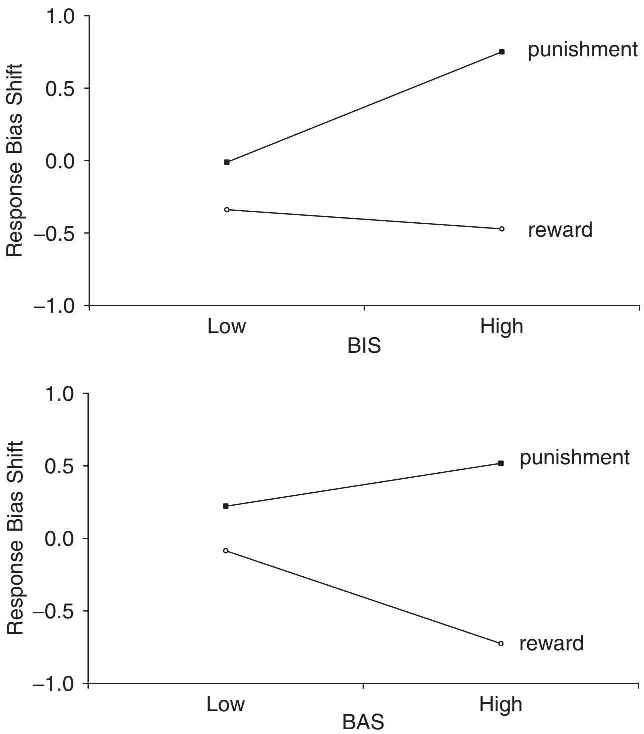

Figure 4 Plot of simple slopes demonstrating an interaction of (a) CWBIS and (b) CWBAS with condition in the prediction of shifts in motivation to respond "Yes" (in terms of the standardized shift in response-bias from a neutral baseline).

NOTE: Response-Bias Shift ($Z_{\Delta}\beta$) is standardized and, as such, zero on the Y-axis corresponds to the mean shift in beta, which was Z = –.92 (reflecting a modest, *p* = .26, overall shift toward responding "Yes"). Low and high values are defined as 1 *SD* below and above the mean, respectively. CWBIS = Carver and White Behavioral Inhibition System; CWBAS = Carver and White Behavioral Activation System.

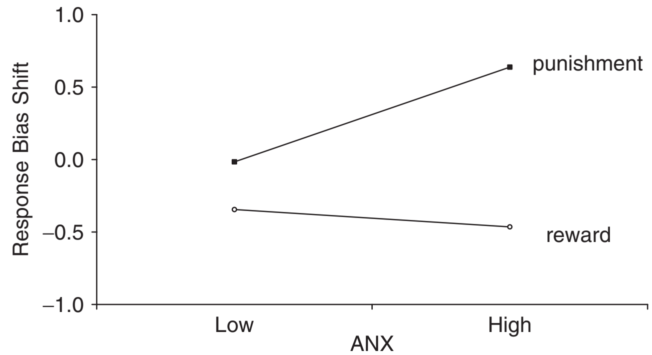

Figure 5 Plot of simple slopes demonstrating an interaction of EPP-Anxiety with condition in the prediction of shifts in motivation to respond "Yes" (in terms of the standardized shift in response-bias from a neutral baseline).

NOTE: Response-Bias Shift ($Z_{\Delta}\beta$) is standardized and, as such, zero on the Y-axis corresponds to the mean shift in beta. Low and high values are defined as 1 *SD* below and above the mean, respectively. EPP = Eysenck Personality Profiler; ANX = Anxiety.

As for Experiment 1, a second MMR examined the generality of relationships observed for the BIS/BAS scales to generic measures of trait impulsivity and anxiety. Main effects (IMP, ANX, and condition) were entered at Step 1 and interaction terms (IMP × Condition, ANX × Condition) were entered at Step 2. At Step 1, the model was statistically significant, R^2 = .14, *F*(3, 77) = 5.16, *p* = .009, with only condition contributing significantly to prediction, B = .72, *t*(77) = 3.23, *p* = .002, as shown by the previous analyses. At Step 2, the increment in prediction fell short of significance, $R^2\,ch$ = .05, *Fch*(2, 54) = 2.40, *p* = .09, and the full model accounted for 19% of the variance in response-bias shifts, R^2 = .19, *F*(5, 54) = 3.55, *p* = .006. Again, condition was a significant predictor, B = .72, *t*(75) = 3.27, *p* = .002, and the ANX × Condition interaction approached significance, B = .39, *t*(77) = 1.74, *p* = .07. Follow-up simple slopes analysis indicated that EPP-Anxiety predicted a significant shift in response-bias away from "Yes" when it was punished, B = .33, *t*(75) = 2.16, *p* = .03, but was not related to response-bias when it was rewarded, B = .04, *t* < 1, *ns* (see Figure 5).

As for Experiment 1, a more lenient test of hypotheses was explored for key relationships that were nonsignificant, in this case to allow for the possibility of a relationship between IMP and response-bias shifts that was not independent of the relationship between ANX and learning. A post hoc MMR analysis (Step 1 = IMP, condition; Step 2 = IMP × Condition) verified that there was no relationship between EPP-Impulsiveness and response-bias shifts, either as a main effect, B = –.13, *t* < 1, *ns*, or in interaction with reinforcement condition, B = –.24, *t*(77) = 1.39, *p* = .17.

Discussion

Main findings from this experiment are as follows. First, Carver and White's (1994) BIS/BAS scales correlated with motivation criteria in a Yes/No categorization task as predicted by Gray's RST. BAS-reactivity as measured by the CWBAS scale predicted development of a response-bias toward a rewarded choice (Hypothesis 3), whereas BIS-reactivity as measured by the CWBIS scale predicted development of a response-bias away from a choice for which responses would constitute goal conflict (i.e., approaching a punishment; Hypothesis 4). These results are consistent with Gray's (1987) arguments relating the BAS as a mediator of approach motivation and the BIS as a mediator of (passive) avoidance motivation. Second, EPP-Anxiety also predicted development of a response-bias that would minimize goal conflict, although this relationship was slightly weaker than that observed for CWBIS. As was concluded from Study 1, this result suggests that trait anxiety may relate to functioning of the BIS, but perhaps not as strongly as a purpose-built BIS scale. Third, EEP-Impulsiveness did not predict response bias, either as a main effect or in interaction with experimental condition. As for the results of Experiment 1, this finding does not support Gray's suggested relationship between BAS-reactivity and trait impulsivity.

GENERAL DISCUSSION

Results from this research provide broad support for two key processes in RST and highlight important issues to be addressed in future investigations. Experiment 1 was designed to evaluate the roles of the BIS and BAS in reinforcement learning, as predicted by stimulus-directed attention and increases in information processing thought to result from BIS or BAS activation. Focal hypotheses were well supported using a purpose-built RST questionnaire, only weakly supported using a generic measure of trait anxiety, and not at all supported using a generic measure of trait impulsivity. An unexpected negative relationship between impulsivity and learning may lend credence to views that link this trait with dysfunctional or maladaptive behavior. Experiment 2 was designed to assess the supposed effect of BIS- and BAS-reactivity on motivation. When the BIS is activated by goal conflict, behavioral inhibition is thought to translate into a motivation to passively avoid punishment, and when the BAS is activated, behavioral activation is supposed to translate into a motivation to approach reward. As for Experiment 1, these predictions were confirmed using purpose-built measures of BIS and BAS. Predictions also were supported using a trait anxiety measure, but again, not for a measure of trait impulsivity. Together, these results have important implications for distinguishing between key processes relevant to RST, devising flexible experimental paradigms for testing predictions from the model, and identifying the traits to which the BIS and BAS correspond.

Learning, Motivation, and a Possible Paradigm

As Pickering (in press) notes, "RST has always been a complex theory, although it masquerades as a simple one" (p. 6). Simplistic or imprecise interpretations of RST, along with less-than-ideal experimentation, have caused a great deal of confusion in the literature, from ambiguity regarding what RST actually predicts to insensitive or even confounded experimental research (e.g., see Pickering & Gray, 2001, pp. 124-132). With this general problem in mind, we have attempted to make explicit two central tenets of RST and derive specific tests to evaluate both. To our knowledge, this is the first experimental program that has attempted to distinguish between behavioral outputs of the BIS and BAS in terms of motivation and learning. These processes are not entirely unrelated aspects of reinforcement but it is likely that many behavior criteria are more related to one of these than the other. Motivation effects may emerge yet remain undetected due to use of a criterion that more closely reflects learning, or both effects could potentially act in opposition of each other (a hypothetical example of which was given earlier). Such possibilities may partially explain some of the mixed findings that have emerged in previous tests of RST.

There are potentially numerous paradigms that might be employed to investigate the two mechanisms highlighted in this article, and some may argue for alternatives to the one that we have adopted. Nevertheless, we believe that certain advantages of a Signal Detection Theory (SDT) approach should not be understated. First, because the parameters of SDT are mathematically orthogonal, they provide distinct quantifications of behavior, reflecting conceptually distinct processes. As such, our experimental criteria do not suffer from the same ambiguity as reaction time or hit-rate (the proportion of correct "Yes" responses), which reflect a combination of learning/performance and/or motivation (Vickers, 1979). (The fact that response-sensitivity and response-bias were examined in separate experiments also is important because these parameters are influenced by qualitatively different task parameters.) Second, although the learning component of the RST appears, at face value, relatively straightforward to test (i.e., by examining increases in any performance measure throughout trials/blocks), the motivational component of RST seems considerably more elusive. Many measures that might be thought to reflect motivation could be equally related to ability (again, reaction time is perhaps a prime example of this). It is therefore convenient that SDT response-bias is strongly linked to motivational factors in that it reflects an individual's preference for one direction of response over the other (McNicol, 1972). Of course, it is not the case that response-bias only reflects motivation, whereas response-sensitivity only reflects learning. Rather, changes in response-bias are principally influenced by motivational factors, and changes in response-sensitivity are principally influenced by factors concerned with the learning of the task.

The fidelity with which measures of response-bias and response-sensitivity represent motivation and learning processes remains to be investigated. Nevertheless, the demonstrated success of the present methodology for testing distinct aspects of RST points to the possibility of paradigm in categorization tasks using an SDT framework.

Chiefly, such a paradigm offers an examination of actual behavior, which seems most appropriate for a model that was, after all, based on ethological research. The clear relevance of behavioral methods to RST was recognized at the outset, as indicated by numerous landmark studies in conditioning (e.g., H. J. Eysenck & Levey, 1972; Nicholson & Gray, 1972; Seunath, 1975). However, these experiments were possibly more suited to examination of the learning component of RST than the motivation component. Motivational processes may be particularly relevant to personality research because these concern the direction of behavior, preferences (rather than abilities), or typical (rather than maximal) responding—notions that all resonate strongly with the view of personality as an individual's typical mode of response (e.g., Pervin, 1990). SDT response-bias offers personality researchers a criterion reflecting motivation that is independent of ability or learning. As such, we see great promise in further explorations of the present methodology in the RST and wider personality literature.

Is Impulsivity the BAS Related Trait?

A secondary focus of the present research was the correspondence among measures of BIS- and BAS-reactivity, such as the BIS/BAS scales and more generic impulsivity and anxiety scales. In the context of personality research, the linking of Gray's biobehavioral systems to trait anxiety and impulsivity are a central tenet of RST. However, the distinction between trait impulsivity and BAS-reactivity has received increasing emphasis (Quilty & Oakman, 2004; Smillie & Jackson, 2006; Zelenski & Larsen, 1999). Consistent with this research, both experiments in the present article showed greater communality among BIS-reactivity measures than among BAS-reactivity measures. Specifically, in Experiment 1, I^7 Impulsiveness predicted poorer learning overall and was not related to superior learning under rewarding conditions, as would be expected if this trait reflected BAS-reactivity. In Experiment 2, EPP-Impulsiveness had no relationship with motivationally based shifts in response-bias in reaction to rewarding feedback.

Although it is therefore tempting to conclude that impulsivity is not a viable candidate for the BAS-related trait, as others have argued, there are some remaining difficulties in disentangling these constructs. First, there is the typically substantial descriptive overlap between BAS/impulsivity scales evidenced by zero-order correlations (Tables 1 & 2). Second, it should be noted that impulsivity measures are, on occasion, found to predict BAS-related criteria relatively well (e.g., Jackson, 2001; Zinbarg & Mohlman, 1998). Third, purpose-built measures of BAS often fail to predict BAS-related criteria, such as some of the Carver and White (1994) BAS subscales in this article (see endnotes 2 and 4). Future research is needed to address this issue, either through employing more sophisticated analyses (see Smillie, Jackson, & Dalgleish, 2006, for a potential solution via multivariate regression) or by targeting a level of analysis (i.e., other than descriptive psychometric level) at which the two constructs may show less descriptive overlap. If BAS-reactivity and trait impulsivity have a reliable relationship, then given the mounting evidence that they are distinct constructs, an explanation for this overlap is surely required.

Limitations and Challenges for Future Research

In RST there are no gold standards for assessing the reactivities of the BIS and the BAS. Some research employs specialized RST questionnaires, whereas other research employs more generic trait impulsivity and anxiety measures—none have yet been clearly distinguished as more valid than the alternatives (Pickering, 2004). In recognition of this problem, we have used both purpose-built RST questionnaires and generic impulsivity/anxiety scales, yet it is unclear to what extent our findings would generalize to alternative measures. It is possible that research employing other purpose-built RST questionnaires will not replicate the results we have obtained using Carver and White's (1994) BIS/BAS scales or that other trait impulsivity measures will be more successful than we found to be the case for I^7 and EPP-Impulsiveness. Replication of the present findings, using multiple measures of RST, is therefore essential (Smillie & Jackson, 2006, provide very similar results to Study 2 using alternative measures of BAS). A longer term goal, furthermore, is the development of more direct measures of BIS- and BAS-reactivity than the current psychometric approach achieves.

A second limitation of these experiments concerns evidence for the physiological changes occurring from BIS and BAS activation, which are thought to underlie the behavioral effects observed. In this research, we chose to focus exclusively on the behavioral level of RST because this is the level with which the primary predictions of RST are concerned. As such, we did not assess physiological changes, such as arousal, that are thought to drive such behavioral effects. Our manipulation checks were able to provide some evidence for effects on affect (Study 1) and behavioral activation/inhibition (Study 2), for instance, but we did not examine changes in arousal or attention. This was because our manipulations were expected to similarly influence arousal and attention but have opposing influences on affect. It would therefore be valuable for future research to determine whether our observed behavioral effects are indeed associated with the physiological processes that theoretically underlie them.

Conclusion

This research has drawn a distinction between two processes in behavioral reactions to reinforcing stimuli within J. A. Gray's model for personality. From this, we derived predictions concerning the relationship between BIS- and BAS-reactivity with reinforcement learning, on one hand, and motivation, on the other. A categorization paradigm was then employed to test these predictions during two experiments. A purpose-built measure of BIS and generic measures of trait anxiety moderated responses to goal conflict, whereas a purpose-built measure of BAS moderated responses to reward. Generic measures of impulsivity did not moderate responses to reward, perhaps supporting the view that such measures do not describe the trait manifestation of the BAS. The possibility of a paradigm based on the present methodology is suggested as a means to overcome one of the most significant challenges for RST: difficulties in deriving clear predictions and providing appropriate tests of Gray's model.

NOTES

1. It must be acknowledged that the Carver and White Behavioral Inhibition System (CWBIS) scale was designed to measure the BIS in terms of Gray's (1982) earlier theory, that is, as a pure punishment system. It is not known to what extent it captures Gray and McNaughton's (2000) revised concept of the BIS as a goal conflict system. The revision of many Reinforcement Sensitivity Theory (RST) questionnaires may well be a necessary agenda for future research (see Smillie, Pickering, & Jackson, 2006).

2. From ancillary analyses, coefficients for the interaction between condition and each of Carver and White's Behavioral Activation System (CWBAS) subscales (Reward-Responsiveness, Fun-Seeking, and Drive) suggested a similar relationship with learning to that of the total scale; however, only that for Fun-Seeking approached significance, $B(3, 56) = .46$, $p = .08$.

3. Also, in the first block, none of the personality variables were correlated with response-bias. The largest correlation was between Eysenck Personality Profiler (EPP)–Impulsiveness and response-bias, $r = .14$, $t < 1$, *ns*.

4. Ancillary analyses showed that two of the CWBAS subscales interacted with condition in the same way as the CWBAS total scale. These were Reward-Responsiveness, $B = .68$, $t(77) = 3.1$, $p = .003$, and Fun-Seeking, $B = .57$, $t(77) = 2.58$, $p = .01$. The interaction between Drive and condition did not approach significance, $B = .20$, $t < 1$, *ns*.

REFERENCES

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. In J. T. Spence & C. E. Izard (Eds.), *Motivation, emotion, and personality* (pp. 137-146). New York: Elsevier Science.

Carver, C. S., & White, T. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319-333.

Cohen, J., Cohen, P., Aiken, L. S., & West, S. G. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). London: Lawrence Erlbaum.

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

Corr, P. J. (2004). Reinforcement Sensitivity Theory and personality. *Neuroscience and Biobehavioral Reviews*, *28*, 317-332.

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

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

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

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., & Levey, A. (1972). Conditioning, introversion-extraversion and the strength of the nervous system. In V. D. Nebylitsn & J. A. Gray (Eds.), *Biological bases of individual behavior* (pp. 206-220). London: Academic Press.

Eysenck, H. J., & Wilson, G. D. (2000). *The Eysenck Personality Profiler* (Version 6). Worthing, UK: Psi-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. (2006). Jeffrey Gray's contributions to theories of anxiety, personality, and psychopathology. In T. Canli (Ed.), *Biology of personality and individual differences*. London: Guilford.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity. *Psychometrika*, *32*, 25-33.

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behavior 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. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford, UK: Oxford University Press.

Gray, J. A. (1987). *The psychology of fear and stress* (2nd ed.). New York: Cambridge University Press.

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

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

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression* (Sage University paper series on Quantitative Applications in the Social Sciences). Newbury Park, CA: Sage.

Jackson, C. J. (2001). Comparison between Eysenck & Gray's models of personality in the prediction of motivational work criteria. *Personality and Individual Differences*, *31*, 129-144.

Jackson, C. J., Furnham, A., Forde, L., & Cotter, T. (2000). The structure of the Eysenck Personality Profiler. *British Journal of Psychology*, *91*, 223-239.

Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, *63*, 1377-1388.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, MA: Cambridge University Press.

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*, 1459-1482.

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrica*, *35*, 237-243.

Mathias, C. W., & Stanford, M. S. (2003). Impulsiveness and arousal: Heart rate under conditions of rest and challenge in healthy males. *Personality and Individual Differences*, *35*, 355-371.

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

McNicol, D. (1972). *A primer of signal detection theory*. London: Allen and Unwin.

Nicholson, J. N., & Gray, J. A. (1972). Peak shift, behavioral contrast and stimulus generalization as related to personality and development in children. *British Journal of Psychology*, *63*, 47-62.

Oei, T. P., Evans, L., & Crook, G. (1990). Utility and validity of the STAI with anxiety disorder patients. *British Journal of Clinical Psychology*, *29*, 429-432.

Pervin, L. E. (Eds.). (1990). *Handbook of personality theory and research*. New York: Guilford.

Petrides, K. V., Jackson, C. J., Furnham, A., & Levine, S. Z. (2003). Exploring issues of personality measurement and structure through the development of a short form of the Eysenck Personality Profiler. *Journal of Personality Assessment*, *81*, 271-280.

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

Pickering, A. D. (2004). The neuropsychology of impulsive antisocial sensation seeking personality traits: From dopamine to hippocampal function? To appear in R. M. Stelmack (Ed.), *On the psychobiology of personality: Essays in honour of Marvin Zuckerman*. New York: 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, UK: Cambridge University Press.

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* (pp. 36-67). Amsterdam: Pergamon/Elsevier Science.

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 individual differences research* (pp. 113-149). Lengerich, Germany: PABST Science.

Quilty, L. C., & Oakman, J. M. (2004). The assessment of Behavioral Activation: The relationship between impulsivity and behavior activation. *Personality and Individual Differences*, *37*, 429-442.

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

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

Smillie, L. D., & Dalgleish, L. I. (2001). Judgement and decision making performance in personnel selection: Partial feedback and the role of confidence. *Australian Journal of Psychology* (Supp.), *53*, 175.

Smillie, L. D., & Jackson, C. J. (2006). Functional impulsivity and Reinforcement Sensitivity Theory. *Journal of Personality*, *74*, 47-83.

Smillie, L. D., Jackson, C. J., & Dalgleish, L. I. (2006). Conceptual distinctions among Carver and White's (1994) BAS scales: A reward-reactivity versus trait impulsivity perspective. *Personality and Individual Differences*, *40*, 1039-1050.

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The New Reinforcement Sensitivity Theory: Implications for psychometric measurement. *Personality and Social Psychology Review*, *10*, 320-335.

Smith, J. E. K. (1969). A note on the standard error of the estimates of the TSD parameters from "Yes-No" experiments. *Michigan Mathematical Psychology Program (MMPP)*, pp. 69-80.

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

Stanislaw, H., & Toderov, N. (1999). Calculation of Signal Detection Theory measures. *Behavioral Research Methods*, *Instruments*, *and Computers*, *31*, 137-149.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301-340.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

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

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

Received November 28, 2005
Revision accepted September 11, 2006