PERGAMON

Personality and Individual Differences 26 (1999) 357–365

PERSONALITY AND INDIVIDUAL DIFFERENCES

Interactions and reinforcement sensitivity theory: A theoretical analysis of Rusting and Larsen (1997)

A. D. Pickering[a,*], P. J. Corr[b], J. A. Gray[c]

[a]*Department of Psychology, St George's Hospital Medical School, University of London, Cranmer Terrace, Tooting, London SW17 0RE, U.K.*

[b]*Department of Psychology, Goldsmiths College, University of London, New Cross, London, U.K.*

[c]*Department of Psychology, Institute of Psychiatry, University of London, Camberwell, London SE5 8AF, U.K.*

Received 4 August 1997

Abstract

In a recent paper Rusting and Larsen (1997) [Rusting, C. L. & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to negative affect: A test of two theoretical models. *Personality and Individual Differences*, *22*, 607–612.] studied the relationship between positive/negative mood induction and personality traits (extraversion, *E*; neuroticism, *N*). They showed that positive mood induction was positively correlated with *E* and negative mood induction was positively correlated with *N* and negatively correlated with *E*. In neither case did an $N*E$ interaction term explain any variance in mood induction over and above that explained by the trait measures entered separately. They argued that the absence of this interaction term was critical evidence for the failure of Gray's theory of *E* and *N* and concluded that these data would fit better within Eysenck's original theoretical framework. This commentary argues that these conclusions are based on an incorrect account of Gray's theory. The correct interpretation is that individuals' susceptibilities to punishment and reward signals (e.g. mood induction cues) are predicted to be *additive* combinations of *E* and *N*, so that statistical interactions between *E* and *N* in explaining mood induction are not expected under this theory. The reasons why Rusting and Larsen may have mistakenly placed so much emphasis on $N*E$ interactions are considered.

1. Introduction

For the last three decades, one of us (J. A. G.) has advocated an account of human personality in which anxiety and impulsivity are proposed to be fundamental dimensions of temperament (Gray, 1970, 1981, 1987; for a review of the recent empirical support for this theory from our own

*Corresponding author.

0191-8869/98 $—see front matter
PII: S0191-8869(98)00019-1

laboratory, see Pickering et al., 1997). Individual differences along these dimensions are argued to reflect variation in the reactivity, or sensitivity, of two basic brain systems to their specific classes of input stimuli. These systems, the behavioural inhibition system (BIS; for anxiety) and behavioural activation system (BAS; for impulsivity), respond to differing kinds of secondary reinforcing stimuli: the BIS is activated by conditioned stimuli signalling punishment or frustrative nonreward (along with innately fear-provoking or novel stimuli); the BAS is activated by conditioned stimuli signalling reward or relief from punishment. In recent years we have referred to this account of personality as reinforcement sensitivity theory (RST).

In a recent article (Pickering et al., 1997) we suggested that RST was actually a much more complex and specific theory than might at first appear and argued that most published investigations into this theory, including our own work, were probably neither adequate nor complete tests of this theory. This impression has been reinforced in a recent article by Rusting and Larsen (1997) which attempted to use a mood induction methodology to compare RST with the related theoretical account of Eysenck (1967) and Eysenck and Eysenck (1985). In this analysis of their work we will show that their key claim about how to test RST is wrong. In addition, with a correct understanding of RST, their claim that the results of their study fit better with the Eysenck model than with RST, can also be seen to be in error.

2. The relationship between RST and Eysenck's theory

From the first article in which RST was espoused (Gray, 1970), the relationship with Eysenck's theory has been emphasised. The anxiety dimension of RST runs (high to low) between the neurotic introvert and stable extravert quadrants of the personality "plane" defined by Eysenck's extraversion (*E*) and neuroticism (*N*) dimensions. Impulsivity is orthogonal to this, i.e. running from neurotic extraverts (high impulsivity) to stable introverts (low impulsivity).

Owing to the explicit theoretical linkage between RST and Eysenck's model, RST has most often been tested, as with Rusting and Larsen (1997), via studies in which the personality groupings/measures have been based on combinations of scores on instruments such as the Eysenck Personality Questionnaire (EPQ; see Pickering et al., 1995, for examples). It is therefore critical to be clear about the precise combinations of *E* and *N* that are predicted (by RST) to relate to BAS- and BIS-mediated behaviours. This section reiterates the predicted combinations, as Rusting and Larsen employ erroneous combinations in their recent paper.

Two versions of RST have been described; once again these versions can be traced back to the initial article (Gray, 1970). The versions differ in the precise relationship with Eysenck's account.

The simplified two-factor account is used for purposes of exposition and illustration and proposes that the anxiety and impulsivity dimensions of RST lie at 45° to Eysenck's *E* and *N* (see Fig. 1). This means that *E* is seen as "a derived dimension reflecting the balance between the reactivities of the behavioural inhibition and approach systems" and *N*, similarly, would reflect "the summation of the reactivities of the behavioural inhibition and approach systems" (Gray, 1987, p. 184).

Using statements concerning proportionality ($\propto$ = "is proportional to") rather than using scaling constants, one can convert the above quotations into a simple, explicit algebra:

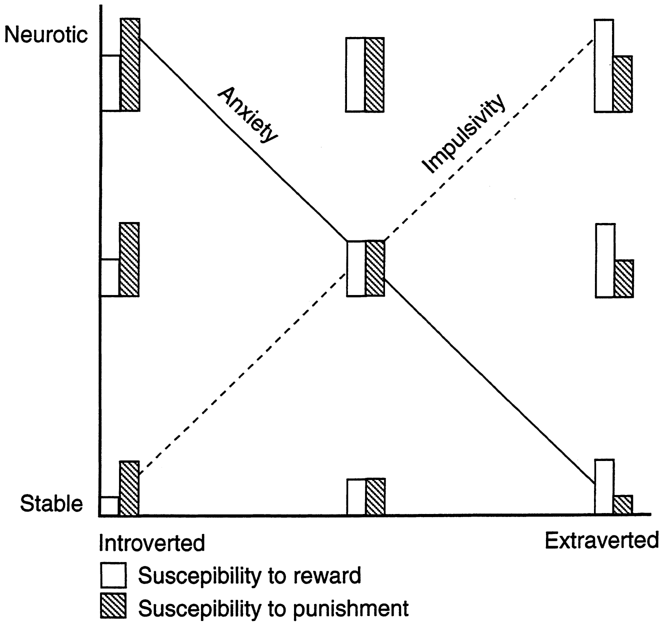

Fig. 1. A graphical representation of the simplified two-factor version of reinforcement sensitivity theory in which the fundamental personality axes (anxiety; impulsivity) are shown at a 45° angle to Eysenck's dimensions of extraversion (horizontal axis) and neuroticism (vertical axis). Although this figure is similar to one which has appeared in previous papers it includes an important correction (see text for details).

$$E \propto (\mathrm{BAS} - \mathrm{BIS}), \qquad N \propto (\mathrm{BAS} + \mathrm{BIS})$$

where the terms represent the values on the relevant personality dimension (the BAS personality dimension being impulsivity; BIS anxiety). These equations can be re-arranged to give the following:

$$\mathrm{BAS} \propto (E+N), \qquad \mathrm{BIS} \propto (N-E) \tag{1}$$

The more realistic two-factor account (see Fig. 2) acknowledges that anxiety lies nearer to N than to E in the $E \times N$ plane. Based on the relative loadings of E and N on anxiety questionnaires, Gray (1970) proposed that the angle between anxiety and N was likely to be closer to 30° than 45°; preserving the orthogonality between anxiety and impulsivity, the angle between impulsivity and E must also be 30°. In terms of the earlier algebra this means that we should really write:

$$\mathrm{BAS} \propto (k \cdot E+N), \qquad \mathrm{BIS} \propto (k \cdot N-E) \tag{2}$$

where k is a weighting constant which would need to take a value of 2 to generate the 30° angles described above.

Now Eqn (1), or more realistically Eqn (2), describe the predicted regressions of Eysenck's E and N on the personality dimensions of RST. They also describe the predicted regressions of E and N on any behaviour exclusively controlled by the BIS or BAS. The beta weights for the E and

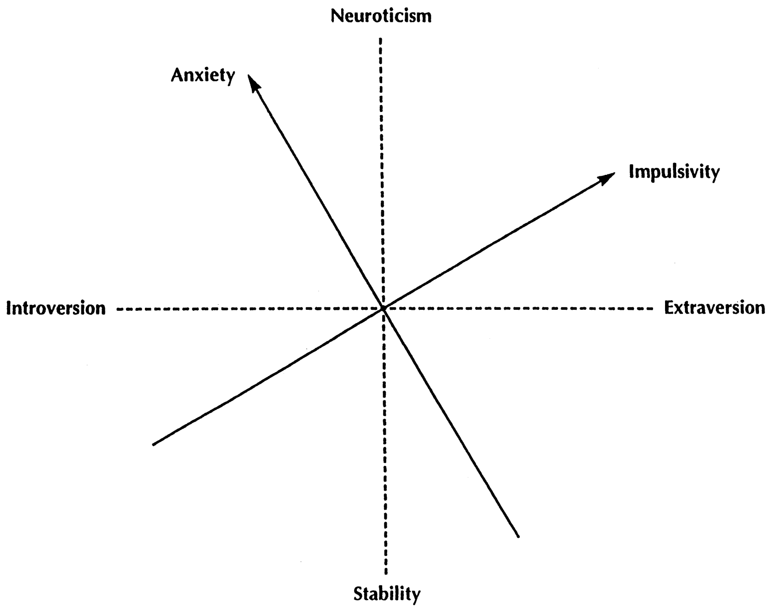

Fig. 2. A simplified geometrical representation of a more realistic set of relationships between the "rival" personality axis systems of anxiety/impulsivity and extraversion/neuroticism. The axes lie at a 30° angle to one another, rather than the 45° angle shown in Fig. 1. The axis labels are at the high poles of the anxiety and impulsivity axes.

N terms, obtained by regressing E and N on BIS- (or BAS-) mediated behaviour, are predicted to be in the k:1 ratio implied by Eqn (2). The actual values of the beta weights cannot be predicted.

3. The Rusting and Larsen (1997) study

Rusting and Larsen's methodology, as noted, involved mood-induction. They argued, quite reasonably, that RST predicts that the extent of positive mood induction (PMI) in a subject should reflect the reactivity/sensitivity of the subject's BAS to secondary positive reinforcers (in this case, the cues used in the positive mood-induction procedure). Similarly, negative mood induction (NMI) should reflect BIS reactivity/sensitivity to secondary negative reinforcers (the negative mood induction cues).

Rusting and Larsen derived, using Eysenck's theory, the following predicted regressions of E and N on the degree of positive (or negative) mood induction:

$$\mathrm{PMI} \propto E, \qquad \mathrm{NMI} \propto N \tag{3}$$

where PMI and NMI refer to the degree of positive and negative mood induction, respectively, and an unknown scaling into actual beta-weights has once again been ignored by using statements of proportionality. Unfortunately, they argued for an inappropriate set of regressions derived from RST:

$$\mathrm{PMI} \propto E + N + c \cdot (N * E), \qquad \mathrm{NMI} \propto N - E - c \cdot (N * E) \tag{4}$$

where c is a constant reflecting the contribution of the $(N*E)$ interaction term relative to the simple effects of N and E. It is clear that Eqn (4), used by Rusting and Larsen, and Eqn (2) which we have always advocated, differ by the addition of the interaction term. Rusting and Larsen repeatedly stressed the importance of this interaction term for testing RST [e.g. "Gray...asserts that there should be an *interaction* between extraversion and neuroticism in predicting positive and negative affect" (Rusting and Larsen, 1997, p. 607, their emphasis)]. This is a fundamental error of understanding RST. We shall consider, below, why they may have been misled into making this suggestion. Before doing so, we shall briefly review and reinterpret Rusting and Larsen's findings in the light of the correct predictions (Eqn (2)).

Rusting and Larsen found (in step 2 of their hierarchical multiple regression), after covarying out baseline levels of positive mood, that E was a highly significant predictor of the positive mood scores of subjects after they had undergone positive mood induction. The beta-weight for E was $+0.46$, approximately twice that for N ($+0.21$), although N did not quite reach statistical significance as an independent predictor of positive mood scores. We have already argued that the results of the subsequent forced-entry of an $N*E$ interaction term (step 3 of their analysis) is irrelevant to the predictions of RST. The fit of these results either to Eqn (3) (Eysenck) or to Eqn (2) (correct predictions from RST) is therefore quite good.

Rusting and Larsen also found, after covarying out baseline levels of negative mood, that N was a highly significant predictor of the negative mood scores of subjects after they had undergone negative mood induction. The beta-weight for N was $+0.49$, approximately 1.6 times the size of that for E (-0.30) and E made a significant independent contribution to the prediction of negative mood scores. (Once again the results of the subsequent forced-entry of an $N*E$ interaction term is irrelevant to the predictions of RST.) Here the results fit the correct predictions of RST (Eqn (2)) very well; rather better, in fact, than they fit the predictions from Eysenck's model which cannot account for the contribution of E to negative affect. We would therefore strongly take issue with Rusting and Larsen's two main conclusions: that the Eysenck model was the better supported in relation to both positive and negative affect susceptibility.

4. Interactions in RST

For psychologists, interactions are probably more familiar in analysis of variance than in multiple regression designs. We shall begin this section by offering one further demonstration that interactions between E and N are not expected under RST. Table 1(A) is formed by taking subjects from the four quadrants of Eysenck's $E \times N$ plane, and shows the relative size of BAS output, for a given BAS-activating input, predicted for each cell by simplified two-factor RST. Table 1(B) is the analogous table for BIS output.

For BAS output, one can see clearly that the effect of N ($+N$ vs $-N$) is a $+1$ change of output for extraverts ($+E$) and $+1$ change for introverts ($-E$). Similarly, the effect of E ($+E$ vs $-E$) is $+1$ change for neurotic subjects ($+N$) and $+1$ for stable subjects ($-N$). Thus the effects of N and E on BAS output are both independent of the level of the other personality factor, which means that there is no interaction between E and N. An exactly analogous argument can be made in relation to BIS outputs using Table 1(B).

Why, then, did Rusting and Larsen (1997) argue for the importance of the $N*E$ interaction so strongly? Part of the reason appears to stem from the fact the authors attempt to lump together the ideas of RST with those of Newman. These two theoretical approaches, although closely related, are different. Specifically, Newman (e.g. Newman et al., 1985; Patterson et al., 1987) has suggested that extraverts and psychopaths (high BAS subjects) have exaggerated responses to reward. It is argued that such subjects form dominant response sets for approach behaviours and that the response set is not readily interrupted by punishment cues. This contrasts with the behaviour of introverts, who show BIS-like interruptions of behaviour in response to punishment cues. Extraverts therefore become increasingly aroused, with a resulting invigoration of any ongoing goal-direct behaviour, when they experience punishments in an approach-behaviour context. This leads, also, to a failure in processing punishment cues effectively. The low sensitivity of extraverts to punishment cues is therefore a specific feature of RST but an indirect consequence of exaggerated responses to rewards in Newman's model. Zinbarg and Revelle (1989) also consider the differences in the predictions of these two theories. Newman's account does not propose an interaction between N and E in determining behaviour; rather the theory stresses the interactions between the signals of reward and punishment in a given situation.

It could be argued that Rusting and Larsen's paper is concerned mainly with Wallace et al.'s (1991) interpretation of RST, in which $N*E$ interactions are relevant (see below). At best, this is an unusual way to investigate a theory, which almost inevitably creates a misleading view of what is under test. We have laid out above, as clearly as possible, the predictions which Rusting and Larsen should have made. Their data, properly interpreted, offer rather good support for RST. The particular extension of RST, which they ascribe to Newman and colleagues, is summarised by the statement that neuroticism is a "gain factor which magnifies or dampens the emotional tendencies due to one's standing on the extraversion–introversion dimension" (Rusting and Larsen, 1997, p. 608). A strict interpretation of "gain", as Rusting and Larsen correctly point out, requires that the predicted regressions include multiplicative $N*E$ interaction terms. The "gain" view of neuroticism appears to be poorly supported by their data, as there is no evidence for any effect of the $N*E$ interaction term on either positive or negative mood induction.

The earlier exposition of RST may also have been partly to blame for suggesting the possibility of $N*E$ interactions. In many of the articles, a particular figure (akin to Fig. 1 here) has been used to illustrate the main points of the simplified two-factor theory (e.g. Gray, 1970, Fig. 7; Gray, 1987, Fig. 10.3). This figure was also used by Wallace et al. (1991) in building their account. The original figures erroneously depict an $N*E$ interaction in the relative heights of the columns showing susceptibilities to signals of rewards and punishments. For example, the height of the punishment susceptibility column for a neurotic subject is shown as 3 times that for a stable subject, irrespective of the subject's position on the extraversion–introversion dimension (the height of the column for the stable subject varying 3:2:1 for introvert:ambivert:extravert). Figure 1 here is redrawn with the appropriate additive relation between N and E.[1]

Rusting and Larsen (1997, p. 611) also claimed that "Newman and colleagues... typically

[1] The error (*mea culpa*; J. A. G.) reveals an interesting result for the interaction model: the level of punishment susceptibility is not constant along the impulsivity axis. An ambivert subject with mid-range neuroticism has a higher punishment susceptibility than either a stable introvert or a neurotic extravert. Similarly these subjects have higher reward suceptibility than subjects at either pole of the anxiety axis (neurotic introverts/stable extraverts). The interaction model cannot therefore be completely squared with the proposed orthogonality between the systems.

find extraversion × neuroticism interactions in predicting sensitivity to signals of reward and punishment''. They listed three empirical studies to illustrate their point (Nichols and Newman, 1986; Bachorowski and Newman, 1990; Wallace and Newman, 1990). Here we review the results of these studies to show that Rusting and Larsen's are incorrect on this point.

In Bachorowski and Newman's (1990) circle-tracing task, the neutral task (simply tracing the circle) produced no significant personality effects at all. In the inhibition task (tracing the circle as slowly as possible) there were main effects of *E*, *N* and task condition (presence vs absence of a specific marked goal for the completion of tracing) but no interaction effects. The authors undertook quadrant analyses and showed that, in the goal condition, neurotic extraverts (high impulsives) were significantly faster than stable introverts (low impulsives).

Wallace and Newman (1990) also used the circle tracing task and compared a goal-only condition with conditions in which a signal of a reward or punishment was present during the tracing. The authors did not report an overall analysis of variance on *E*, *N* and tracing condition, going instead to planned quadrant comparisons between high and low impulsives in the reward and goal conditions (high impulsives were significantly faster). As Figs 1 and 2 (plus Table 1) make clear, the effects found in these quadrant analyses would be expected if there were just main effects of *E* and *N* (as was the case in Bachorowski and Newman, 1990); they do *not* imply an $N*E$ interaction. It might appear that quadrant analyses suggest an $N*E$ interaction when these comparisons reveal larger differences than either of the *E* or *N* main effects alone. This will occur, even without any $N*E$ interaction, because the quadrant analysis looks at the *N* and *E* main effects combined (see Table 1).

Nichols and Newman (1986) carried out two experiments using a pattern-matching task. In experiment 1, the effects of personality on response latencies were explored under three separate conditions: with noncontingent rewards for correct responses; with noncontingent punishments

Table 1
Outputs from the BAS (A) and the BIS (B) as a function of neuroticism (*N*) and extraversion (*E*)

	$+N$	$-N$
A		
$+E$	2	1
$-E$	1	0
B		
$+E$	1	0
$-E$	2	1

+/−: Denote high and low scores, respectively, on the personality dimension concerned. 0 = low output; 1 = medium output; 2 = high output. These hypothetical outputs are in response to appropriate BIS- and BAS-activating stimuli of fixed intensity.

for errors and with both noncontingent reinforcers being used. Once again there were no $N*E$ interaction effects, although data from the high and low impulsive quadrants were presented in a figure. In experiment 2 just the reward and punishment conditions were repeated and planned quadrant analyses were reported without an overall analysis of variance. Thus, in none of the studies cited by Rusting and Larsen were $N*E$ interactions actually found. Rusting and Larsen may have been misled into thinking that such interactions were present by the use of quadrant analyses in these studies.

To close this commentary, we must consider the complications which might possibly result from the fact that RST proposes that the BIS and the BAS mutually inhibit one another. Could this interaction between the systems lead to a statistical interaction between E and N in their effects on BIS and/or BAS mediated behaviour or moods? Recently, Pickering (1997) developed a formal model of BIS and BAS mutual inhibition using neural network techniques. A full description of the model is beyond the scope of this commentary, and interested readers are referred to the original article. However, unpublished explorations using this model have shown that the interaction between the BIS and the BAS would not be likely to produce statistical interactions of the kind discussed by Rusting and Larsen.

In summary, this commentary has stressed several points: that the punishment and reward susceptibilities of RST are predicted to be *additive* combinations of E and N, so that statistical interactions between E, N and reinforcement manipulations are not expected under this theory; that such interactions have not been found as often as is sometimes claimed; and that researchers must remember that the simplified two-factor model (45° rotation) is used to simplify presentation of theory whereas the more realistic model involves combinations of N and E with unequal weightings (e.g. a 30° rotation). Finally, the Rusting and Larsen (1997) results in fact lend rather good support to RST.

References

Bachorowski, J.-A., & Newman, J. P. (1990). Impulsive motor behavior: Effects of personality and goal salience. *Journal of Personality and Social Psychology, 58*, 512–518.

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

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

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

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

Newman, J. P., Widom, C. S., & Nathan, S. (1985). Passive avoidance in syndromes of disinhibition: Psychopathy and extraversion. *Journal of Personality and Social Psychology, 48*, 1316–1327.

Nichols, S. L., & Newman, J. P. (1986). Effects of punishment on response latency in extraverts. *Journal of Personality and Social Psychology, 50*, 624–630.

Patterson, C. M., Kosson, D. S., & Newman, J. P. (1987). Reaction to punishment, reflectivity, and passive avoidance learning in extraverts. *Journal of Personality and Social Psychology, 52*, 565–575.

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

Pickering, A. D., Corr, P. J., Powell, J. H., Kumari, V., Thornton, J., & 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 personality: Tribute to Hans J. Eysenck at eighty*. London: Elsevier Sciences.

Pickering, A. D., Díaz, A., & Gray, J. A. (1995). Personality and reinforcement: An exploration using a maze-learning task. *Personality and Individual Differences, 18*, 541–558.

Rusting, C. L., & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to negative affect: A test of two theoretical models. *Personality and Individual Differences, 22*, 607–612.

Wallace, J. F., & Newman, J. P. (1990). Differential effects of reward and punishment cues on response speed in anxious and impulsive individuals. *Personality and Individual Differences, 11*, 999–1009.

Wallace, J. F., Newman, J. P., & Bachorowski, J.-A. (1991). Failures of response modulation: Impulsive behavior in anxious and impulsive individuals. *Journal of Research in Personality, 25*, 23–44.

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