

STRESS REACTIVITY; RESEARCH TRENDS AND INDIVIDUAL PROGNOSIS

A REPLY TO J. E. TONG

F. J. S. ESHER*, J. E. ORME*, D. W. McKERRACHER**

* Middlewood Hospital, Sheffield 6

** Rampton Hospital, Nottinghamshire

It would appear from J. E. Tong's paper that our article (1965) has perhaps created the impression with some readers that we were attempting to discredit the findings of Tong and Murphy. We made no attempt to attack the theory underlying their work, nor to question the findings of the original authors. We appreciate Tong's point that the original studies were based upon *groups* of patients, specially *selected* for stability and instability, and we accept that they are not to be thought of as representative of the Rampton population at any date.

Our report was an attempt to show whether the techniques could be used *clinically* for prognostic purposes, in dealing with individual patients. There is obviously a difference between research trends found in specially selected groups, and the application of autonomic indices based on these trends in the context of a routine clinical assessment. Since our study was unsuccessful in establishing the tests on this basis, we concluded that, ". no simple replication is possible and to that extent, caution must be exerted in the clinical interpretation of those techniques." Many factors appear to be involved in the production of autonomic response patterns, and we simply felt that it was premature to add such techniques to a standard psychological test battery, without further exploration of their usefulness. We did not depart dramatically from the techniques that Tong and Murphy reported, and such alterations as did occur were no more than might be expected between one psychologist's room and another. If the difference between half-reclining and sitting introduces such differences as Tong suggests, then this only emphasises the delicacy of the technique, and undermines its possible usefulness as a clinical tool.

Reply to particular points made by J. E. Tong

(1) The results of the study by Esher et al appear to involve at least one statistical error

Tong suggests that there are more statistical errors than one in our data without enumerating them. If he has noted any in addition to the one he mentions he should have provided the evidence so that we could have replied. If Table 4 (to which he refers) is consulted, it may be seen that we did not in fact quote any statistical figures for Chi-Square, or any other statistical test, so that it is difficult to see where the mistake in our calculations occurs. The reason for not applying a chi-square was on account of the weak cell (a frequency of only 2) in the Hollow Meadow's data. This renders the chi-square supplied by J. E. Tong of doubtful value though the direction of the results is in line with his theory as he suggests. Perhaps we were wrong to say there was an "absence of *any* association between temperature change and stability rating," but we could not assert the opposite with any conviction, especially since the Rampton results did not confirm the Hollow Meadow's distribution of temperature scores. Our impression was that the overall trend of the results was not sufficiently strong to make any positive agreement possible with the earlier findings.

(2) A reasonably objective social criterion is the only satisfactory basis with which to compare test scores

Longitudinal studies involving extensive follow up of patients after their discharge might appear to provide an empirical means of estimating social stability, but we would question whether such a process could ever be 'reasonably objective'. The usual procedure of allowing each individual an equal time period of exposure to risk, takes no account of the social influences and pressures experienced by each in an unsheltered environment. The varying degrees of stress suffered must surely play a vital role in shaping the ensuing personality reactions. A patient with a high relapse potential may not relapse, because he is fortunate in the small amount of stress he experiences on his return to community living.

We stated explicitly that we had not conducted a follow up after discharge, and we have therefore no comment to offer concerning Tong's interesting findings utilising such a procedure.

(3) From the information given by Esher et al, it appears that their rating of stability bears little relationship to that reported by Tong or Murphy

We disclaim Tong's charge that our stability ratings were dissimilar to his and Murphy's. The method used was exactly the same as that described by Murphy (1961, p. 515). It is not usually considered necessary in an article to go into the details of what was done, when reference is made to the original source of the procedures.

Murphy had two means of establishing criteria of stability for patients:

(a) *Nurses' ratings*: he found that the G.S.R. eye-touch conditioning scores did not discriminate significantly between the stable and the unstable within a hospital environment, but that a similar technique using an eye-puff stressor was successful in doing so at the .05 level of probability. However, this significant finding was only achieved by collating the low and high conditioners, and comparing their score distribution with that of the medium conditioners. Murphy did not therefore substantiate Tong's findings in a straightforward fashion, and there was no clear-cut indication that low reactivity was a definite indication of psychopathic instability as Tong suggests (Murphy 1961, p. 517-18).

(b) *Relapsed patients*: these were composed of a mixture of readmissions to Ramp-ton plus 34 patients who were followed up for a minimum period of 13 months. During this period only three men relapsed. This necessitated the inclusion of readmissions to provide sufficient numbers. Interestingly, the results were reversed on this occasion: the eye-touch conditioning method was successful where the eye-puff method failed. Again the discrimination was at the .05 level of probability. This time, however, the low and high conditioners did not have to be summated to get this result.

| | <i>CRs</i> | | |
|-------------------|------------|-----|------|
| The figures were: | 0-3 | 4-7 | 8-11 |
| Success | 8 | 8 | 16 |
| Relapse | 10 | 3 | 4 |

So far as prediction is concerned, it can be seen that low conditionability does not inevitably imply relapse potential, since 8 of the 18 low responders were in fact successful. It is the highly responsive group which is better related to the prediction of

success in the open community, closely followed by the medium range patients. Thus, low conditionability may or may not be associated with relapse in a prognostic sense, whereas *high* conditionability as a criterion has an 80% chance of selecting the likely to be successful patients.

(4) One would expect that Rampton patients, being extremely unstable by definition would contain more people of low conditioning score than an ordinary open hospital as Hollow Meadows

Tong is at pains to point out that the results he obtained from his 1959 study were based upon cases specially selected as definitely stable or unstable. As he himself points out, these scores are not representative of the whole Rampton population. It is difficult, therefore, to see how he can generalise about the meaning of his conditioning scores from these specially chosen groups to the whole Rampton population of his own day, let alone attempt to generalise to the present Rampton population. One might have made a tentative hypothesis about extreme social misbehaviour and its physiological correlates, but one could not possibly have *predicted* that our findings would turn out as they have done.

We have already pointed out that in the quoted Murphy figures (section 3 (b) above), low conditioners were almost evenly divided between success and failure. On the basis of such results, the actual predictions that might have been made would have been: (a) Hollow Meadows and Rampton would share almost equally a large number of low conditioners. (b) Hollow Meadows would have a significantly greater number of medium or high conditioners.

Our data showed the latter hypothesis to be acceptable, but the former one was not substantiated, since Hollow Meadows had few low conditioners. Our article can therefore hardly be claimed by Tong as "independent support" for his suggestions concerning low scores in this laboratory measure.

(5) Further evidence of the predictive significance of the low score is given by Esher et al by showing that 17 of the 21 low scorers as assessed five or more years earlier are still rated unstable

It is puzzling to know why Tong criticises us for making this statement, since at that point we were in fact suggesting that his procedure was more successful. We assumed that the changes (some of them unavoidable) we had introduced in the testing environment may have had some effect upon the scores: "It is interesting that the earlier conditioning scores should maintain some connection ($P = .1$) with behaviour after such a lapse of time when the recent ones bear none, and does suggest that the failure of the more recent conditioning technique is at least not wholly due to complete alteration in the reactivity of the patients".

If the changes we introduced (having the patient more comfortable in less threatening surrounds) did drastically affect the efficacy of the stress technique, this itself would be an interesting fact. It would suggest that the G.S.R. conditioning technique may not be a specific stressor technique at all. It could be regarded as merely a focal stressor imbedded in the total laboratory situation stresses. Its results might only hold good when the general environment in which it takes place is perceived as threatening by the patient, and a sufficient degree of general stress arousal is achieved. This would certainly explain the failure of the recent Rampton research work to find a link between nurses' ratings and the G.S.R.

(6) The data used for comparison with the Esher et al scores was not comparable

(a) Tong objects to the comparison of our data with that of Murphy. He points out that Murphy's patients were selected for being physically aggressive, and were not representative of the Rampton population. In an article to be published (Comparison of the behaviour problems presented by male and female subnormal offenders: McKerracher, Street and Segal, Brit. J. Psychiat.) it was noted that 82% of the recent male admissions had committed an act of physical violence of some kind prior to admission. This underlines the fact that the main problem presented in a special security establishment in the "sixties" is that of violent aggression. Admittedly Murphy's and McKerracher's samples were not paired for comparison, but they did share aggressive tendencies to a significant degree. If Tong wishes to maintain that neither of these samples can be compared with his in the "fifties", we could certainly agree about our own sample since that was the implication of our published findings!

(b) A direct comparison with Tong's data was not made, because the bulk of his was based on patients resident in Rampton prior to the 1959 Mental Health Act. Murphy's data was recorded during the transitional phase, after the introduction of the Act, but before it became fully operational. It was felt that Murphy's data was therefore, chronologically more relevant.

(c) One of the main conclusions of Murphy's thesis was that his findings resembled those of Tong:

- (i) in the bimodal distribution (U-shaped) of conditioned responses obtained from unstable patients (rated in hospital).
- (ii) in being related to nurses' ratings of stability in the same way as Tong's.

For these reasons it is difficult to see why it is not valid to have used Murphy's data as a basis for comparison.

(d) Tong compares his figures published in 1962, with ours for 1964 (not 1965 as he says) and finds there is no significant difference. In fact, if the figures are split into Low Conditioning Rate v Rest (cut off point <4), then chi-Square=3.042. This falls just short of .05 level of significance, which Tong himself accepts throughout as a reliable indication of a trend.

The difference in size of samples can go a long way towards explaining the lack of significance found when more than a two-cell breakdown is attempted. If we can be allowed to offer further evidence and supply the figures for 1965 which are now to hand, but were not at the time our article went to press, then this point will be seen to be important.

| | | CRs | | | |
|----------------------|-----------|--------|-----|-----|------|
| | | Method | 0-3 | 4-7 | 8-11 |
| Rampton (Tong 1962) | eye-touch | | 142 | 149 | 110 |
| " (McKerracher 1964) | eye-touch | | 38) | 32) | 13) |
| | | |)77 |)62 |)26 |
| " (McKerracher 1965) | eye-puff | | 39) | 30) | 13) |

X^2 for total distribution = 10.449, $P = < .01$
 X^2 for Low CR v Rest = 6.284, $P = < .02$

Contrary to Tong's assertion, there does appear to be a consistent difference between the distribution now, and that prior to the 1959 Act. Far fewer highly responsive patients are now found. Whether this is due to the demographic change, or whether it is due to artefacts produced by our handling of the technique, is not clear.

Incidentally, we would like to point out that we did *not* conclude there had been a demographic change in the Rampton population *because* our scores differed from those of Murphy. That such a change had taken place was demonstrated in the article mentioned above by McKerracher, Street and Segal. Knowing that there was such a difference in the populations, we hypothesised that the G.S.R. scores might *reflect* this alteration at a physiological level.

(7) The inclusion of Mentally Ill and brain-damaged patients in Esher's sample further invalidates their comparison

We are not certain which part of our article Tong is attacking by making this assertion, since we certainly made no statements about including such patients. Naturally one would expect meaningless results if a hotch-potch of cases was assembled in such a fashion as is implied. We state categorically now that only sub-normal and psychopathic patients without known brain lesions, or psychotic symptoms, and not under medication, constituted our sample groups.

(8) There is little similarity between the rating methods of Esher et al and the original studies. Similar comments have been made in connection with the stressors originally used

Tong appears not to believe us when we said that we had attempted to replicate his techniques. We assure him that the instructions, concerning the administration of both stressors, were carried out faithfully, as described in Murphy's doctoral thesis (1961). The changes we mentioned were in the environmental surrounds. The biggest change was to have the patients half-reclining rather than sitting. This should not have affected a *specific* stressor stimulus.

(9) Esher et al make references to Pavlovian theory and arousal theory. They appear to confuse conditioning and arousal

We would like to remind Tong that the choice of words was his. We did not introduce the concept of conditioning into his work. He made reference in all his articles to Conditioning Scores. Naturally we assumed that he was referring to conditioning theory when he used this term. If he did not intend to convey this impression, then he should have avoided using such a term. Furthermore, in his criticism of our article in this journal, having just completed an argument leading to the rejection of conditioning theory as an explanation of the G.S.R. scores, and the election of "stress activity" to that role, he goes on to make this curious contradictory statement:

"The drive function of stress anxiety must be considered for any theory of delinquency using stress variables. Consequently, a conditional or learned response is more directly relevant than a simple stress response".

If this sentence does not refer to conditioning theory, then it must mean that the "learned response" is to be regarded as a *complex* stress response, to differentiate it from a "*simple* stress response". Presumably the latter phrase refers to an unconditioned reaction. Apart from the different names given to the various reactions, we find it difficult to distinguish between Tong's stress reactivity interpretation and the conditioning theory.

The main reason for choosing to use Cr. scores in our study was in order that our data could be compared with that of Murphy. Apart from this, we did replicate Murphy's findings that there were highly significant correlations amongst the preparatory, the "conditioned" and the "unconditioned" responses. It is, there-

fore, permissible to state that these scores are arranging people in much the same order of responsiveness. To this extent, the scores could be viewed as interchangeable. This in no way implies that the scores represent the same aspect of arousal, or conditionability, or stress reactivity, or whatever the G.S.R. is measuring. Our point was that the distinctions may be possible theoretically, in *practice* it would make little difference to selection of patients if unconditioned responses were substituted for "conditioned" ones; or, to a lesser extent, if preparatory responses alone were used. With this in mind, we then questioned whether the concept of conditionability was clearly distinguished from general reactivity, and if not, whether it was *clinically* worthwhile applying a time-consuming technique when simpler autonomic indices might be just as effective. This is not the same thing as attempting to discredit Tong's theory at a theoretical level.

(10) The Esher report highlights the necessity of using adequate criteria of stability in research with psychopathic subjects. Their figures indicate that "stability" has a different meaning, for example, for Hollow Meadows staff than Rampton staff

Instability is a relative term, since all judgments are relative to the environment they occur in. Naturally, if Hollow Meadows' nurses were to rate Rampton patients, they might be expected to consider them all unstable compared even with the worst Hollow Meadows' patients. No attempt to achieve cross-ratings of this nature was made. Nurses' judgments were restricted to their own hospital, and to the expectations of the different disciplinary demands made upon their own patients. If this particular G.S.R. technique is to become a robust clinical test, then it will be required to differentiate sheep from goats in any hospital, not just in Rampton.

(11) The ubiquitous nature of Tong's cut-off points deserves some comment

Some of Tong's published figures refer to "<4 and >3" as the cut-off point for low reactivity, but the 1959 figures that he quotes for us himself show that low reactivity includes scores up to 6 with the cut-off point being placed at <7 and >6. Furthermore, his original theory was based on the idea that there were two different mechanisms underlying instability: stress activation and stress conservation. In his thesis, he summated low and high conditioners under the general heading of instability, and compared this group with patients scoring in the medium range of the conditioning dimension. If this latter procedure was invoked for the data in Table 4 of our article, the breakdown would become:

| | low & high CR | medium CR |
|----------|---------------|-----------|
| Stable | 15 | 15 |
| Unstable | 24 | 18 |

This is by no means significant, and shows how results can be altered by introducing enough permutations. Clinically, it is extremely important to know whether a patient scoring in the medium range is to be considered as a reasonable risk, or regarded as a low conditioner and a relapse potential. When numerous different cut-off points are selected in this fashion, one runs the risk of obtaining a significant result by chance alone, especially when the levels of significance reported are mainly at $P=.05$ level.

(12) The non-significant correlation of +.18 (Pearson r) between Tong's conditioning scores and those quoted by Esher et al on the same patients, falls just beneath the .05 level of probability and suggests a long-term reliability

We would agree with Tong that there is no reason to suppose that the G.S.R. scores are reliable over a long period of time, since it is to be expected that as patients improve, their autonomic reactivity may change. However, he seems to be

suggesting that though patients probably do alter in reactivity the correlation of $+ .18$ also shows there is some continuity which falls just short of significance. In fact this correlation indicates that less than 4% of the variance is in common, and it is meaningless to base any hypothesis upon it alone. Other means of examining this question are more relevant (see next section).

(13) One reason for the correlation of $+ .18$ being so low may be the tendency for some patients to change dramatically from low reactivity to high reactivity and vice versa during the course of their sojourn in Rampton

We were very interested in this possibility pointed out by Tong and re-examined our figures to compare our CR scores with his. The result was as follows:

| | | | | | | | | | | | | |
|---|---|-------|---|--|-------|---|--|-------|---|---|----|----|
| | <i>Patients whose scores remained in same range</i> | | | <i>Patients whose scores moved up or down by one range</i> | | | <i>Patients whose scores changed from one extreme to the other</i> | | | | | |
| N=74: | 23 | (31%) | | 37 | (50%) | | 14 | (19%) | | | | |
| <i>Numerical differences between the CR scores of Tong's group and those of Esher et al</i> | | | | | | | | | | | | |
| N=74 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | 4 | 15 | 9 | 3 | 15 | 2 | 13 | 5 | 6 | 0 | 0 | 2 |
| | | 28 | | | 20 | | | 26 | | | | |
| | | (38%) | | | (27%) | | | (35%) | | | | |

It would seem that around 30% of the patients did *not* change appreciably over a 5 year period, but approximately 20% *did* change dramatically. Half of the patients showed movement away from the original scores in non-specific directions. It is probably the latter scores which have reduced the correlation, rather than the scores of those who changed from low-high reactivity.

From the practical point of view it is interesting to note that almost 40% of the patients scored within 2 points of the original Tong scores. This certainly does raise the possibility for some long-term G.S.R. reliability in *some* patients but it is important to stress that this is of *theoretical* interest only, since from the practical and clinical point of view only 60% of the patients showed marked alterations. This means G.S.R. scores cannot be used predictively over a 5 year span, and that patients can be expected to change considerably during this period.

Further data might be of general interest at this point.

A series of G.S.R. conditioning studies over a period of time was conducted on those patients who had already been tested by Tong, and who were still in Rampton by the time the series was terminated. The N of 74 was reduced to N=40 at the termination of the experiment. Correlations quoted for N=40 were Pearson r:

| | | | | |
|---------------------------|-------------------|-------------------|-------------------|-------------------|
| | <i>1st repeat</i> | <i>2nd repeat</i> | <i>3rd repeat</i> | <i>4th repeat</i> |
| Tong | $+ .195$ | $- .054$ | $+ .079$ | $+ .028$ |
| 1st repeat | | $+ .405^{**}$ | $+ .286$ | $+ .209$ |
| 2nd repeat | | | $+ .593^{**}$ | $+ .451^{**}$ |
| 3rd repeat | | | | $+ .908^{**}$ |
| ** significant at $P=.01$ | | | | |

Tong=CRs obtained 1958-61 using eye-touch conditioning procedure.

1st repeat=CRs obtained 1963 (Feb.-July): 5 years gap: eye-touch conditioning.
 2nd repeat= " " 1963 (Aug.-Jan.): 6 month gap: eye-touch conditioning.
 3rd repeat= " ") 1963-64 (Oct.-Mar.): 2 mth. gap: eye-puff conditioning.
 4th repeat= " ") Carried out on morning and afternoon of same day.

It would seem that there is a gradual and steady reduction in the correlation as the length of time expands. On the same day, with the same technique (eye-puff) the correlation is +.9; with an interval of two months it is between +.4 and +.6, depending on whether identical or merely similar techniques are used after 6 months, it is between +.2 and +.4; after 5 years it is between 0 and +.2. This supports Tong's contention that there is some long-term reliability, but it also suggests that the G.S.R. scores can be relied upon not to change substantially, only over a 6 month period at most. Even then, though the correlation is significant, it would be hazardous to use them for selection purposes. It could be said that over a 2 month period there is reasonable consistency of G.S.R. reactivity.

(14) Esher et al do not supply the data showing the present stability ratings of nurses with their CR scores obtained from the sample with N=74

We did not supply the figures, because we said in the text they were not significant. For the record they were as follows:

| | <i>low cond.</i> | <i>med. cond.</i> | <i>high cond.</i> | Total |
|----------|------------------|-------------------|-------------------|----------|
| Stable | 0-3 20 | 4-7 9 | 8-11 9 | 38 |
| Unstable | 18 | 13 | 5 | 36 |
| | <hr/> 38 | <hr/> 22 | <hr/> 14 | <hr/> 74 |

The skew towards the low end of the dimension is quite marked in both groups. The similarity between this result, and that demonstrated in Table 3 of our original article is remarkable.

CONCLUSIONS

There would appear to be general agreement that subnormal patients can be spread along a dimension of G.S.R. reactivity. We would also agree with Tong that many subnormal patients who are incapable of surviving in the community unsupported reveal themselves to be highly reactive, or markedly unresponsive. The data of Tong and Murphy suggested that both types were formerly to be found in Rampton. However, the data of Esher *et al* suggests that mainly unresponsive patients are currently being admitted to Rampton, and relatively few highly responsive ones. The fact that the Hollow Meadows' sample was chiefly highly reactive is interesting, if it can be deemed representative of the kind of unstable subnormal who is now being directed to local subnormal hospitals.

The point at issue is whether the second study can be said to replicate the findings of the first. If one limits the appraisal to the Rampton patients alone, then it can be shown that nurses' stability ratings bear no relationship to the physiological criteria, and in this respect the study failed to get similar results. Moreover, we still maintain that the distribution of low and high responders differs between now and then. This could either be a physiological fact, or a test artefact, but whatever the reason, it can be said that in this respect too, the second study failed to get the same findings as the first.

If one widens the appraisal to include all subnormals who are unable to survive in the community without getting into trouble of some kind, then it can be said that the recent study was successful in separating between those who were most violent and dangerous (presumably the unresponsive ones at Rampton), and those who were most inadequate and agitated (the highly responsive ones at Hollow

Meadows). This cannot be said to be merely a replicatory finding, since the original authors made no such wide survey. Their research work was concentrated upon Rampton patients alone.

In other words, the second study produced a new fact: namely that as the 1959 Mental Health Act appears to have brought about demographic changes in the Rampton population (see section above), the G.S.R. techniques might now be more useful as a general screening device, *outside* a particular hospital environment. It does not seem capable of discriminating so well *within* a special hospital environment, because the majority of patients there nowadays are unresponsive. It could be said that the general trend of the earlier results was reproduced in a broader spectrum, but not within the narrower context of the original studies.

References

- Esher, F. J. S., Orme, J. E., McKerracher, D. W. (1965). Replicatory Studies of two Psycho-Physiological Techniques of Assessing Mentally Subnormal Patients. 3 *Ment. Subnorm.* XI, 93-98.
- McKerracher, D. W., Street, D. R. K., and Segal, L. J. (1965). A Comparison of the Behaviour Problems Presented by Male and Female Subnormal Offenders. *Brit. J. Psychiat.* in press.
- Murphy, I. C. (1961). *Stress Reactivity and Anti-Social Aggression*. Ph.D. Thesis. Sheffield University Library.
- Tong, I. E. (1962). Psycho-Physiological Studies in Psychopathy and the Prediction of Stability. *Proceedings of the London Conference on the Scientific Study of Mental Deficiency*, Vol. 1.