

PERFORMANCE IN GROUP TELEPATHY EXPERIMENTS AS RELATED TO TARGET PICTURE CHARACTERISTICS

Jan Dalkvist & Joakim Westerlund¹
Department of Psychology, Stockholm University, Sweden

ABSTRACT

A re-analysis of previous data in a project on telepathic group communication of emotions, as evoked by slide pictures, was performed in an attempt to identify picture characteristics that were correlated with performance. Performance, measured as receivers' ability to discriminate between 15 positive and 15 negative target pictures looked at by senders in an adjacent room, was analysed as a function of psychological and psychophysiological properties of the target pictures.

Each of the 30 target pictures was rated by 66 undergraduate students, 24 males and 42 females, on six graphic scales. Four of the six scales measured purely emotional aspects of the pictures, namely how (a) pleasant/unpleasant, involving, (c) compassion-arousing and (d) repulsive they were. The two remaining scales measured (a) how well-known and (b) how easy to apprehend the motifs of the pictures were.

Two studies were performed to measure physiological responses to the target pictures. In both studies, participants were undergraduate students, who were tested in small groups.

Electrodermal activity (EDA) was measured using an EDA monitor, with a software program for collecting and analysing data. Sixty participants, 32 females and 28 males, provided the data used in the study.

Heart-rate (HR) was measured using an electronic HR metre, consisting of two parts: (a) a belt with heart-beat sensors and (b) a "watch" for receiving and storing signals from the heart-beat sensors, which were analysed using an associated software program. Fifty participants, 36 females and 14 males, provided the data used in the study.

Two different data sets were analysed, one from an old study, involving 240 participants, and the other from a new study, involving 605 participants. The two data sets were analysed both separately and together, with mean number of negative guesses as the response measure.

The eight stimulus scales were combined into an index scale of general negative arousal. For participants who started as receivers, the index scale exhibited a significant positive correlation with the mean number of negative guesses for the new data set ($r(29) = .382$; $p < .05$, two-tailed) and for the total one ($r(29) = .393$, $p < .05$, two-tailed). For the same groups of participants, a subsequent correlation analysis involving the eight separate stimulus scales showed the correlation between the involvement scale and mean number of negative guesses to be the strongest one (new data set: $r(29) = .52$, $p < .01$, two-tailed; total data set: $r(29) = .45$, $p < .05$, two-tailed).

¹We would like to acknowledge the financial support of John Björkhem Memorial Foundation.

The question of which stimulus characteristic(s) (if any) determined the participants' responses was discussed. A possible explanation of the lack of positive results in the old data set in terms of geomagnetic disturbances was also discussed.

It was pointed out that the present results confirm previous findings of a sender/receiver order effect. It was also pointed out that assessment of the validity of the present results hinges on results soon to be obtained in an ongoing replication study.

In spite of its practical advantage of being able to provide a large amount of data in a short period of time, group experiments on extrasensory perception (ESP) are relatively rare. One reason is probably the common opinion that group testing is, for some reason, inferior to individual testing in producing positive results (see, e.g., Rhine, 1947/1971, p. 40). There is little empirical support for this view, however. Thus, not only negative results (e.g., Haight, Weiner & Morrison, 1978; Milton & Wiseman, 1999), but also positive results have been obtained using groups of respondents in ESP experiments (Barker, Messer & Drucker, 1975; Carpenter, 1988, 1991; Dalkvist & Westerlund, 1998), just as in the case of individual testing. More importantly, the number of group experiments on ESP that have been reported so far is too small to permit any well-founded evaluation of this type of experiment. In particular, almost no group experiments on telepathy have been reported.

One exception is a series of group telepathy studies that have been performed at the Department of Psychology, Stockholm University, since 1993; these studies were initiated by one of the present authors (JD). Based on the idea that strong emotional messages (e.g., signals of danger) may be, for evolutionary reasons, easier to transmit telepathically than are more neutral messages (Moss & Gengerally, 1969), the studies have all been concerned with transmission of emotions, as evoked by slide pictures.

When the participants (typically about 15 psychology students, who participated in the study as part of course requirements) arrived at the laboratory, they were randomly divided into two groups, one sender group and one receiver group. The senders and the receivers were sequestered in two soundproof rooms, with one room in between. The two experimental rooms were connected to each other by a signal device: a lamp in the receiver room could be turned on and off from the sender room.

The stimuli used were 30 slide pictures, 15 with positive motifs (such as nature pictures and pictures of happy people) and 15 with negative ones (such as pictures of traffic accidents and starving children). The slides were presented in random orders, a new order for each group of senders. The senders' only task was to look at the pictures and to "hold on to" the feelings evoked by the respective pictures as long as they were shown. The receivers were instructed to guess whether a given picture was positive or negative. One of the (two) experimenters in the receiver room (there were also two experimenters in the sender room) watched the signal lamp and reported to the receivers when a new picture was being shown to the senders. Each picture was shown for 20 seconds, with an inter-stimulus interval of about half a second.

When all 30 pictures had been shown, the participants changed rooms, and those who had served as senders in the first part of the experiment now served as receivers and vice versa.

Hit rate, defined as number of correct responses or proportion of correct responses, was invariably used as the dependent variable in the data analyses. Hit-rate was analysed as (a) a function of person or situational factors (e.g., belief in telepathy and the order in which the

participant served as sender and receiver and (b) a function of stimulus factors (e.g., rated characteristics of the target pictures).

As a first part of the present project (Dalkvist & Westerlund, 1998), five individual studies, which mainly served to generate a set of hypotheses, were conducted. These hypotheses were subsequently tested in a comprehensive replication study (Westerlund & Dalkvist, 2004). The outcome of this study was negative: none of the eight predictions was borne out. A new finding, however, concerned with the order of the roles of sending and receiving telepathic messages, was obtained. Specifically, respondents starting as receivers showed a significantly higher hit-rate than did respondents starting as senders – an effect that in a subsequent analysis turned out to be correlated with disturbances in the earth's magnetic field (Dalkvist & Westerlund, 2006). An additional re-analysis (Dalkvist, Montgomery, Montgomery & Westerlund, 2008), concerned with variability in performance rather than mean performance, seemed to confirm the existence of a sender/receiver order effect: It showed highly significant reduced variability in mean performance among groups of respondents with the same sender/receiver order.

In order to obtain a psychological description of the pictures used, they were rated on six different scales (by participants other than those taking part in the main experiments). Initially, all of the pictures were rated with respect to how (a) pleasant-unpleasant, (b) involving, (c) familiar and (d) perceptible (easy to apprehend) they were. In order to get a more exact description of the negative pictures with respect to evoking emotional reactions, they were also rated with respect to how (a) compassion-arousing and (b) repulsive they were, using a new group of raters. In addition, the 15 positive pictures were rated with respect to how calm or exciting they were, using still another group of raters. Moreover, on the basis of the four "emotional" scales, a global scale of negative emotionality was created.

Although this global scale was predicted to be negatively related to hit-rate, it failed to show any significant such relationship in the follow-up study (Westerlund & Dalkvist, 2004). Our method of testing for effects of stimulus characteristics was not optimal, however, because positive and negative stimuli were tested separately, to avoid a response bias caused by preference for one of the two response alternatives.

When at the start of our present project, hit-rate was chosen as the dependent variable, we assumed, quite naturally, that the receivers would follow the instructions and discriminate between positive and negative pictures, if they were able to report any telepathic information at all. On closer reflection, however, this assumption is not at all obvious. After all, emotional reactions are complex phenomena, varying in at least two global dimensions: pleasure-displeasure and arousal (perceived and/or physiological), the latter dimension exhibiting a more or less pronounced U-formed relation to the former (Lang, Greenwald, Bradley & Ham, 1993). Thus, the notion that some form of arousal rather than pleasure-displeasure (or any more specific pleasurable or displeasurable experience) was transmitted from the senders to the receivers seemed conceivable and worth testing.

Increased electrodermal activity (EDA) and decreased (!) heart rate (HR) enter as pivotal components in a pattern of various physiological reactions, known as the orienting response, evoked by motivationally relevant stimuli, capturing the individual's attention (Sokolov, 1960). The two reactions are not perfectly correlated, however. While both of them tend to be somewhat stronger in response to negative stimuli as compared to positive ones, the

difference between negative and positive stimuli is somewhat more pronounced for the decrement in HR than for the increase in EDA (Lang et al., 1993).

The major purpose of the present study was to re-analyse our previous data in terms of target picture characteristics using both subjective stimulus scales (new ones, to be described below) and two physiological scales (HR and EDA) to see if, and how strongly, they were related to participants' ability to discriminate among our target pictures.

TARGET PICTURES

Subjective and physiological properties of the 30 target pictures were measured in three different studies. New subjective measures were collected from participants in some of the experiments in the above-mentioned replication study (Westerlund & Dalkvist, 2004), immediately after the main experiments were finished. This was done to obtain more reliable scales than the initial ones, by using a greater number of raters than before (about 15 raters per picture were then used) and the same raters for all scales.

Two studies were performed to measure physiological responses. In the first study, the target pictures were measured with respect to EDA and in the second study with respect to HR.

The target pictures (for a detailed description of the pictures, see Dalkvist & Westerlund, 1998) were projected on a white screen or a white wall (the EDA study). The target pictures were shown in a randomized order. The projector, located behind the participants, was run automatically, by means of a timer.

Subjective measurements

Participants

Participants were 66 undergraduate students at Stockholm University, 42 females (mean age: 27.2 yrs) and 24 males (mean age: 27.4 years). Most of the participants were psychology students who chose to participate in the study as part of course requirements. All participants were informed beforehand that some of the pictures to be shown were very disgusting, and sensitive persons were recommended not to take part in the experiment.

Procedure

The ratings were made in groups, with ten or fewer participants in each group.

Each picture was rated on six different scales. Four of them measured purely emotional aspects of the pictures, namely how (a) pleasant/unpleasant, (b) involving, (c) compassion-arousing and (d) repulsive they were. The two remaining scales measured (a) how well-known and (b) how perceptible (easy to apprehend) the motifs of the pictures were.

The participants were given a booklet of forms, each one containing the six scales, with one form for each of the 30 target pictures. For half of the subjects in a given session, the scales were written in the above order; for the other half, the order was reversed. Similarly, for half of the sessions, the target pictures were presented in one particular order and for the other half in the reversed order. Each target picture was presented for 30 sec. The judgments were made by drawing a vertical line on a 100-mm-long graphic scale with verbally anchored endpoints,

representing extreme states of the experience to be judged (for example, “not at all disgusting” and “very highly disgusting”).

Final scales

For each scale and target picture, the arithmetical mean was calculated over all judges.

The inter-rater reliability was assessed for each scale individually using Cronbach’s alpha. The mean alpha was 0.985 (*Std. Dev.* = 0.016).

EDA measurements

Participants.

Participants were 77 undergraduate students at the University of Stockholm, 43 females and 31 males, most of them at the Department of Psychology, who had chosen to participate in the study as part of course requirements.

Due to disturbances in the equipment or loose electrodes, only 60 participants, 32 females and 28 males, entered into the final sample. These participants varied in age from 19 to 41 yrs., with a mean age of 24.72 yrs. All participants were informed beforehand that some of the pictures to be shown were very disgusting, and sensitive persons were recommended not to take part in the experiment.

Apparatus.

An EDA monitor, with a software program for collecting and analysing data, manufactured by Biopac Systems inc., was used. The monitor (model mp 100A) and the program (AcqKnowledge III for the 100WS version 3.2) were linked to and installed in a Macintosh 6320 computer.

The electrodes used were of the EL204S-Ag/AgCL type. Isotonic paste served as electrolyte (0.5% NaCl/100ml H₂O). The electrodes were fastened to the middle phalanxes of the middle finger and the ring finger of the nondominant hand.

Procedure.

At most three participants took part in each measurement session. Each picture was presented for 20 sec., with an inter-stimulus interval of about half a second. The total number of sessions was 35. Each session lasted for about 20 min., including instructions.

Three participant chairs, with elbow-rests, were located in a row in front of the projector.

The sequence of events during a session was as follows:

1. The participant(s) were asked to wash their hands with soap and water.
2. The participant(s) were given a complete description of the experiment, except for any expectation of the results or the connection to previous telepathy experiments. They were then informed that they were allowed to close their eyes if a picture was experienced as too unpleasant, and even to break off the experiment.
3. The experimenter instructed the participant(s) on how to attach the electrodes by doing it on her own fingers and then letting the subject(s) do it themselves (without electrode paste).

4. The experimenter attached the electrode paste to the electrodes and the participant(s) fastened the electrodes.
5. The participant(s) were instructed (a) to keep the hand with the electrodes as still as possible, but without concentrating too much on it, and (b) to put their arms as comfortably as possible on the arms of the chair. The participant(s) were also told not to speak to each other or to the experimenter during the experiment.
6. The light in the experimental room was turned off and the projector was started. The first picture (a photo of a landscape) was not entered among the thirty "real" stimuli, but was only used to accustom the participant(s) to the experimental situation (without the participant(s) knowing it).
7. After the last trial, the participant(s) were disconnected from the monitoring equipment and asked to fill out a short questionnaire, mainly concerned with demographic data.
8. The participant(s) were debriefed with respect to the purposes of the study. (In addition to testing the telepathy hypothesis, the data were also used in a study on gender differences in emotion.)

Final scale.

For each target picture, the mean of the response amplitude was calculated over a 5 sec. interval, starting from the exposure of the picture. These means were then transformed into z -values for each participant separately, to eliminate effects of individual response level differences. Finally, for each of the 30 target pictures, the mean z -value was calculated across the sixty (nonexcluded) participants.

As measured by Cronbach's alpha, the inter-respondent reliability was 0.511.

HR measurements

Participants.

Participants were 53 undergraduate students at the University of Stockholm, 37 females and 16 males, most of them at the Department of Psychology, who chose to participate in the study as part of course requirements. Due to missing data, only 50 participants, 36 females and 14 males, entered into the final sample (only occasional missing data were accepted). These participants varied in age from 19 to 39 yrs., with a mean age of 26.80 yrs. As before, all participants were informed beforehand that some of the pictures to be shown were very disgusting, and sensitive persons were recommended not to take part in the experiment.

Apparatus.

HR was measured by means of an electronic HR metre, manufactured by Polar Electro Oy. The HR metre (S610iTM) consists of two parts: (1) a belt with heart-beat sensors, to be fastened around the chest, with the sensors adhering to the skin just below the breasts or the breast muscles, and (b) a "watch" for receiving and storing signals from the heart-beat sensors. Associated with the HR metre is a software program for data analysis.

Procedure.

Four participants at most took part in each measurement session. Each picture was presented for 20 sec., with an interstimulus interval of about half a second. Each session lasted for about 20 min., including instructions. There were 19 measurement sessions in all.

In order to avoid interference among the HR monitors, participants were spread out in the experimental room a couple of metres apart. In order to prevent females from becoming embarrassed when attaching the sensor belts, males were located in front of the females.

On the experimenter's commands, the participant(s) started their "watches" just before the start of the experiment and stopped them when the experiment was finished. Also, again on the experimenter's commands (saying "now"), the participant(s) pressed a button to register the point in time when a picture was exposed, and another button 7 sec later to delineate the time interval for the measurement (the minimum time interval required for HR calculations was 5 sec., that is, the same period as that used in measuring EDA, but with that short period we would have run the risk of obtaining periods that were too short to be possible to analyse, due to delayed button presses when a picture was exposed.) The participant(s) could hear a "beep" when the button had been properly pressed.

The sequence of events during a session was as follows:

1. The participant(s) were seated on chairs as described above.
2. The participant(s) were given a complete description of the experiment, except for any expectation of the results or the connection to previous telepathy experiments. They were then informed that they were allowed to close their eyes if a picture was experienced as too unpleasant, and even to break off the experiment.
3. Following the experimenter's instructions, the participant(s) fastened the sensor belts as described above.
4. The participant(s) were told not to speak to each other or to the experimenter during the measurements and to concentrate on the pictures and not on handling the watch.
5. The participants were instructed how and when to start and stop their watches and how and when to make the time registrations, as described above.
6. A short trial run, without any pictures, was carried out.
7. The light of the experimental room was turned off, the participants started their "watches", the experimenter started the projector, and so on.
8. After the last trial, the subjects took off their sensor belts and "watches" and filled out a short questionnaire, mainly concerned with demographic data.
9. The participants were debriefed with respect to the purpose of the experiment.

Finally, the data were transmitted to a computer for analysis.

Final scale

By means of the program mentioned above, for each target picture, the mean response amplitude was calculated over a 7 sec. interval, starting from the exposure of the target picture. These means were then transformed into z -values for each participant separately, to eliminate effects of individual differences in response level. Finally, for each of the 30 stimulus pictures, the mean z -value was calculated across the fifty (nonexcluded) subjects.

As measured by Cronbach's alpha, the inter-respondent reliability was 0.240.

Scale comparisons

As expected from previous work (Lang et al., 1993), the EDA and HR scales were found to be negatively correlated. The correlation was small and nonsignificant ($r = -0.17$; $p = 0.358$), however.

Table 1 shows the Pearson correlations among the six rating scales. As can be seen from this table, most of the correlations are strong or very strong. As can be further seen, the four emotional scales (1-4) are positively correlated with each other, but negatively correlated with the two, positively correlated, nonemotional scales (5-6).

Table 1. Pearson correlations among the six subjective scales

Scales	(1)	(2)	(3)	(4)	(5)	(6)
Pleasure-Displeasure (1)	1					
Compassion (2)	.96**	1				
Repulsion (3)	.98**	.97**	1			
Involvement (4)	.71**	.85**	.77**	1		
Well-known (5)	-.74**	-.62**	-.75**	-.32	1	
Perceptible (6)	-.53**	-.40*	-.53**	-.03	.81**	1

Significant at the 2% level, two-tailed

Table 2 shows the Pearson correlations between the two physiological scales and the six rating scales. As can be seen from this table, all of the four emotional rating scales are positively correlated with the EDA scale and negatively correlated with the HR scale, while the two nonemotional rating scales exhibit the reverse correlation pattern. It may also be noted that the correlations obtained for the HR scale are stronger than those obtained for the EDA scale, inspite of the fact that Cronbach's alpha is higher for the latter scale than it is for the former one.

Table 2. Pearson correlations between the two physiological scales and the six subjective scales

Subjective scales	Physiological scales	
	EDA	HR
Pleasure-Displeasure	.38*	-.47**
Compassion	.36	-.43*
Repulsion	.39*	-.49**
Involvement	.30	-.42*
Well-known	-.31	.54**
Perceptible	-.15	.32

*Significant at the 5% level, two-tailed
**Significant at the 1% level, two-tailed

DATA ANALYSIS

The same data as those used in the above-mentioned study on the sender/receiver order effect (Dalkvist & Westerlund, 2006) were used in the present study. Calculations were made for the old and the new study separately as well as for both studies combined.

As indicated in the introduction, a single experiment consisted of two sessions (= runs), one in which the participants served as senders and the other in which they served as receivers. The

respective distributions of participants, sessions and experiments for the old, the new and the total data set are given in Table 3.

Table 3. Distribution of participants, sessions and experiments over three data sets

Unit	Data		
	Old ^a	New	Total
Participant	240	605	845
Session	34	90	124 ^b
Experiment	17	47	64

a. The last three out of five studies

b. Four sessions were discarded because of technical failures

The data were also divided into two subsets based on sender/receiver order.

Each of the six rating scales and each of the two physiological scales was correlated with the mean number of negative guesses across the 30 target pictures. The mean number of negative guesses was chosen as the response measure instead of mean hit-rate (which had been used in previous studies) to allow calculation of correlations across both positive and negative stimulus targets simultaneously without the correlations being affected by a response bias. That is, a tendency to choose one of the two alternative responses over the other would lead to a higher hit rate for the preferred response alternative than for the nonpreferred one. When mean hit-rate had been used as the response measure in previous studies, correlations could only be calculated across positive and negative pictures separately to avoid response bias. No such problem occurs when the mean number of negative guesses, or a corresponding measure for positive guesses, is used. This is because, on the null hypothesis, these measures become randomly distributed among both positive and negative target pictures independent of any tendency to prefer one of the two alternative responses.

In running group experiments, one is most often confronted with a particular statistical problem, called the “stacking” problem: due to the possible occurrence of dependency among participants’ responses in group testing (for example, caused by the occurrence of a common response bias, such as a tendency to give one type of response at the beginning of a run and another type at the end of it), the statistical assumption of independent measures runs the risk of being violated, leading to deflated *p*-values (Thouless & Briar, 1970). In the present study, however, there was no stacking problem, because the degrees of freedom were determined by the number of stimuli and not by the number of respondents.

RESULTS

Given the strong correlations among the eight stimulus scales, all of them were combined into an index scale, for an initial global analysis. This was done by (1) reversing the signs of the three scales that were negatively correlated with the other five scales, namely the HR scale, the familiarity scale and the perceptibility scale, (2) transforming the values of each scale into *z*-values, and (3) summing all eight scales. This index scale can be conceived of as a general scale of negative emotional arousal.

Table 4 shows the Pearson correlations between the index scale and the mean number of negative guesses for both sender/receiver orders taken together and for each sender/receiver order separately for the old, the new and the total data set. As can be inferred from this table,

in the new and the combined data set, there was a significant positive correlation for those participants who started as receivers, while participants who started as senders exhibited nonsignificant negative correlations. No significant correlation was obtained for the old data set.

Table 4. Pearson correlations between the index scale and the mean number of negative guesses for both sender/receiver orders taken together and for each sender/receiver order separately for the old, the new and the total data set

Data set	Both sender/receiver orders	Receiver first	Receiver last
Old	.109	.104	.051
New	.051	.382*	-.273
Total	.103	.393*	-.104

*Significant at the 5% level, two-tailed

More specific results are given in Table 5, showing the Pearson correlations between the mean number of negative guesses and the eight stimulus scales for both sender/receiver orders taken together and for each sender/receiver order separately for the total data set. As can be seen from this table, for subjects who started as receivers, all scales exhibited positive correlations (ranging from .33 to .35) except for the HR, the familiarity and the perceptibility scales, significant correlations being obtained for compassion and involvement, the strongest one for involvement.

Table 5. Pearson correlations between stimulus scales and mean number of negative guesses for both sender/receiver orders taken together and each sender/receiver order separately

Stimulus scales	Both sender/receiver orders	Receiver first	Receiver last
EDA	.07	.36	-.16
HR	-.24	-.25	.07
Pleasure-Displeasure	.02	.33	-.02
Repulsion	.06	.35	-.10
Compassion	.22	.37*	-.12
Involvement	.04	.45*	-.10
Familiarity	.00	-.21	.14
Perceptibility	-.09	-.10	-.01

*Significant at the 5% level, two-tailed

As can further be seen in Table 5, for the subjects who started as senders, the correlations showed a tendency to have opposite signs to the correlations for subjects who started as receivers. There was no statistically significant correlation in this condition, however, most of the correlations falling close to zero.

As can be seen from Table 6, the pattern of correlations shown in Table 5 for the whole data set was strengthened in the new data set, but is totally absent in the old one.

Table 6. Pearson correlations between stimulus scales and mean number of negative guesses for both sender/receiver orders taken together and for each sender/receiver order separately for the old and the new data

Stimulus scale	Old data			New data		
	Both Sender/receiver orders	Receiver first	Receiver last	Both Sender/receiver orders	Receiver first	Receiver last
EDA	-.19	.02	-.29	.21	.40*	-.09
HR	-.06	-.12	.03	-.24	-.21	-.13
Pleasure-displeasure	.17	.10	.14	-.09	.31	-.39*
Repulsion	.14	.06	.15	-.02	.36	-.35
Compassion	.12	.06	.12	-.01	.38*	-.35
Involvement	.03	.01	.03	.25	.52**	-.15
Familiarity	-.09	-.06	-.06	.10	-.19	.30
Perceptibility	-.24	-.22	-.13	.16	.03	.18

*Significant at the 5% level, two-tailed

**Significant at the 1% level, two-tailed

DISCUSSION

Only the participants who started as receivers and came from the new study showed consistently significant positive results. These participants showed a significant correlation between the index scale, which can be interpreted in terms of general negative arousal, and mean number of negative guesses. It should be borne in mind, however, that neither the participants in the new study who started as senders, nor any of the participants in the old study showed any significant correlation between the index scale and mean number of negative guesses. This may be taken as a warning that our statistically significant results could be due to multiple testing, even though the number of tests involving the index scale was small.

The strongest correlation was obtained for the involvement scale. Although this specific finding could be due to sampling errors increasing the correlation for the involvement scale as compared to the correlations for other scales and therefore should not be taken too seriously, it does suggest that the participants' responses were based not on the pleasure-displeasure dimension per se, but rather on emotional involvement or some related picture property (if there was any effect of picture characteristics at all). One possibility is that the participants' responses were based on emotional intensity independent of whether the emotion was positive or negative. Alternatively, the responses were basically determined by some form of physiological arousal, in spite of the fact that the involvement scale was more strongly related to the mean number of negative guesses than was the EDA or the HR scale. If so, the weaker results for the EDA and the HR scale as compared to the involvement scale could be due to the fact that the former scales have lower reliability than does the latter one. Likewise, the fact that the EDA scale exhibited a stronger correlation with the mean number of negative guesses than did the HR scale could be because the former scale is more reliable than the latter one.

Unfortunately, it is impossible to draw any firm conclusion based on the present results regarding the relative merits of subjective and physiological stimulus correlates in research on

telepathic transfer of emotions, as evoked by external stimuli. The major reason is the relatively low reliability of the two physiological scales, particularly that of the HR scale.

The fact that the two sender/receiver orders did not yield the same results is consistent with previous analyses of the present data, showing a significant difference between the mean hit-rate of the two sender/receiver orders, with a higher hit-rate for groups of participants starting as receivers than for groups of participants starting as senders (Dalkvist & Westerlund, 2006). In that study, while participants who started as receivers tended to perform above chance level, those who started as senders tended to perform *below* chance level. Not surprisingly, a corresponding tendency was obtained for the pleasure-displeasure scale and for the repulsion and the compassion scale in the present study. However, for the involvement scale as well as the two arousal scales, only those participants who started as receivers showed any deviation from chance level. This apparent difference between different stimulus scales should be followed up in future research.

As in previous studies (Dalkvist & Westerlund, 2006; Dalkvist et al., 2008), there was a difference between the old and the new data set, only the new one giving positive results. In these studies, the difference between the old and the new data set could, apparently, be related to lower levels of geomagnetic disturbances in the new study than in the old one. Considering that the same data were analysed in the present study as in the previous ones, this tentative explanation should be applicable to the present study as well, even though no test of the explanation in question was made here.

Once again, analyses have revealed a difference between the two sender/receiver orders as far as the new data set is concerned. Thus, the possibility that there is a true difference between the two sender/receiver orders has been strengthened in the present study.

The most critical question is, of course, whether or not the present results can be replicated. Fortunately, we will soon get the answer to this and other questions concerning the validity of the present data in our ongoing replication study, involving another set of approximately six hundred participants.

REFERENCES

- Barker, P.L., Messer, E., & Drucker, S.A. (1975). Intentionally-Deployed Attention States: Relaxation. A Group Majority Vote Procedure with Percipient Optimization. *Proceedings of the Parapsychological Association 16th Annual Convention*, 165-167.
- Carpenter, J.C. (1988). Quasi-therapeutic group process and ESP. *Journal of Parapsychology*, 52, 279-304.
- Carpenter, J.C. (1991). Prediction of forced-choice ESP performance: III. Three attempts to retrieve coded information using mood reports and a repeated-guessing technique. *Journal of Parapsychology*, 55, 227-280.
- Dalkvist, J. & Westerlund, J. (1998). Five experiments on telepathic communication of emotions. *The Journal of Parapsychology*, 62, 219 - 253.
- Dalkvist, J. & Westerlund, J. (2006). Telepathic group communication of emotions: Announcement of predictions for an ongoing experiment. *Proceedings of Presented Papers: Parapsychological Association 49th Annual Convention*, 314-319.
- Dalkvist, J., Montgomery, W., Montgomery, W., & Westerlund, J. (2008). Re-analysis of group with focus on variability. *Submitted*.

- Haight, J., Weiner, D., & Morrison, M. (1978). Group testing for ESP: a novel approach to the combined use of individual and shared targets. *Proceedings of the Parapsychological Association 24th Annual Convention*, 96-98.
- Lang, P.J., Greenwald, M.K., Bradely, M.M., & Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 261-273.
- Milton, J., & Wiseman, R. (1999). A meta-analysis of mass-media tests of extrasensory perception. *British Journal of Psychology*, 90, 235 – 240.
- Rhine, J.B. (1947/1971) *The reach of the Mind*. New York: William Sloane.
- Sokolov, E.N. (1960). Neuronal models and the orienting reflex. In M.A. Brazier (Ed.), *The central nervous system and behavior*. New York: Macy.
- Thouless, R.H., & Brier, R.M. (1970). The stacking effect and methods of correcting for it. *Journal of Parapsychology*, 34, 124-128.
- Westerlund, J., & Dalkvist, J. (2004). A test of predictions from five studies on telepathic group communication of emotions. *Proceedings of the Parapsychological Association 47th Annual Convention*, 269-277.