Anomalous baseline effects in mainstream emotion research using psychophysiological variables

Dick J. Bierman
University of Amsterdam, The Netherlands
and
University of Utrecht, The Netherlands

The goal of this study was to see whether anomalies observed in physiological baseline measurements could be found in data from previously published studies. Three datasets were reanalyzed. The first dataset was from a study on the speed with which fear arises in animal phobic participants vs. controls. The second study was concerned with the difference in anticipatory responses prior to choosing cards from risky vs. non-risky decks of cards in a gambling task. The third dataset was from a study investigating the effect of emotional priming on the evaluation of Japanese characters.

In all three studies marginally significant anomalous effects were found. The anomaly was an unexpected difference in the baselines preceding randomized emotional vs. calm stimuli. In the studies using emotional pictures this was mostly due to the baselines preceding erotic pictures. The combined result across the three studies was significant ($z = 2.748$, $p = 0.003$).

A secondary goal of this study was to estimate, independent of the reason for the anomalous baseline effect, how much this effect might have influenced measurement of the response values. It is shown that the error introduced by taking a baseline value at or just before the start of the stimulus ranges from 10% to 30% of the main effect, at least in the three studies investigated here. This finding carries implications for the interpretation of results in mainstream psychophysiological research.

Keywords: Skin Conductance baselines, Emotion, Presentiment, Intuition

---

1. Introduction

In psychophysiological research on emotions, subjects are typically presented with emotional or calm stimuli while psychophysiological measures are continuously monitored. The dependent variable in studies of this kind is generally the post-stimulus response to the aforementioned stimuli. In this article we will restrict ourselves to skin-conductance measures as the dependent variable of interest. The response value of this variable can be operationalized in many ways. One could use the peak value obtained in a specified time period following the stimulus, or the integral over the signal during this period, etc. (Boucsein, 1992). In the majority of studies a baseline value, usually derived from the signal just before presentation of the stimulus, is subtracted. The reason for this baseline correction is to reduce variance in the signal that is not related to the stimulus *per se*, but to previous exposures or to habituation to the target sequence.

In a study where subjects were presented with a randomized sequence of emotional and calm visual stimuli, Radin (1997) found suggestive evidence that the baseline values appeared to be related to the subsequent stimuli. Most notably the baseline level of skin conductance preceding highly emotional stimuli was higher than the baseline level preceding calm stimuli. This anomaly appeared to be experimentally robust, but also rather surprising, because if the effect was real it would imply that a commonly used procedure for calculating response values in psychophysiological studies was biased in an unknown way.

To determine whether these anomalous baseline effects might have been caused by a problem with the instrumentation, we decided to replicate the original study with entirely different hardware and software. The results were basically similar, albeit slightly smaller in magnitude, in three independent experiments (Bierman & Radin, in press). Given the possible relevance for mainstream emotion research, we reported these findings as a note in *Perceptual and Motor Skills*, concluding that the reported effects might indicate that commonly employed double-blind, randomized protocols still allowed subjects to (unconsciously) infer the category of the forthcoming stimulus (Bierman & Radin, 1997).

The questions asked in the current investigation were: 1) Are these apparently anomalous baseline effects also present in paradigms that are not specifically designed to measure them, and 2) if so, how serious an error would this constitute in the calculation of the response values or other normal variables of interest?

To answer these questions we reanalyzed data from three previously published psychophysiological experiments by researchers who were unaware of this anomaly, and with completely different research goals in mind. The advantage of using these independent datasets is that they can shed light directly on the prevalence and magnitude of the effect in generally accepted research paradigms where randomization procedures are used to prevent the subject from out-guessing the category of the upcoming stimuli. The disadvantage is that the experiments were not intended to exclude all normal explanations of a possible anomalous baseline effect, including inappropriate randomization. As a consequence, we cannot draw generalized conclusions about the

reality of the anomaly, but we can identify whether these anomalous baseline differences appear in "unsuspected" data. Thus, if the effect is found, it may stimulate other researchers to investigate this issue, and in the process either find a normal explanation or establish a true anomaly.

2. Global description of datasets used in the reanalysis.

Three datasets from three different paradigms in psychophysiological emotion research were selected for reanalysis. The necessary data was retrieved for each of these three experiments, all of which used strong emotional stimuli and skin conductance as the dependent measure.

The first dataset was obtained by request from Prof. Alfons Hamm (of the University of Greifswald, Germany). The data were skin conductance samples from an experiment exploring the speed with which fear arises in animal-phobic subjects after a picture with the fear-inducing animal is shown. The data for the control group were also made available. The experiment was reported by members of Hamm's research group (Globisch et al, 1999). The setup and especially the timing of the stimuli were close to the experimental setup used in the original studies suggesting the anomalous baseline effect (Radin, 1997; Bierman & Radin, 1997). We refer to this study as the "animal-fear study."

The second dataset was obtained from graphs published by Prof. Antonio Damasio's group from the University of Iowa Medical School, in *Cognition* and *Cerebral Cortex* (Bechara et al, 1994, 1996). These data concerned the skin conductance of brain-damaged and normal subjects while they participated in a gambling task. Specifically, skin conductance was measured just before subjects took a winning or losing card from one of four randomized decks of cards. These decks were designed to be more or less advantageous in the long run. The goal of the study was to investigate if subjects' physiology reflected learned, unconscious knowledge about the decks before the subjects were consciously aware that the decks were biased. This study differed considerably from the studies originally suggesting the anomalous baseline differences. Most notably, the emotional event was not induced by a well timed pictorial stimulus but by a less well-controlled feedback of the sum of money that was won or lost. We refer to this experiment as the "gambling study."

The third dataset was obtained from a master thesis experiment at the University of Amsterdam investigating the effect of emotional primes on the evaluation of Japanese characters (Durieux, 1999). During this experiment, which was essentially a replication of a study by Murphy and Zajonc (1993), skin conductance was measured because certain theoretical frameworks (e.g., LeDoux, 1996) suggested that the conscious evaluation of the Japanese character is driven by non-conscious processes which might be reflected in the subject's physiology. The experimenter provided the raw dataset directly to the present author. We refer to this study as the "emotional priming" study.

For all three studies the global hypothesis was that the baseline (anticipatory skin conductance) preceding emotional events would be greater than baselines preceding non- or less-emotional events. For details of the studies that are not relevant for the understanding and evaluation of our reanalysis, refer to the original publications.

3. The animal fear study [1]

3.1 Participants

Eighty six participants (54 women; 32 men; ages 18-41) were selected from a student population. Subjects were included in the high animal fear group if they scored above the 85th percentile of the distribution in their gender group on either a spider or a snake fear questionnaire. Volunteers were assigned to the control group if their scores fell below the 50th percentile on either of these questionnaires.

3.2 Procedure & materials

Each participant viewed 60 color slides, in part selected from the International Affective Picture system (IAPS, Center for Study of Emotion and Attention, 1995). From these 60 slides, twenty were snakes or spiders, depending on the scores on the animal fear questionnaires, 8 were erotic pictures, and 32 were calm pictures like mushrooms, household, animals or flowers. For 38 subjects (14 animal fearful) the pictures were presented for 150 msec preceded by a fore-period of 7 seconds during which a fixation point was shown (see Figure 1).

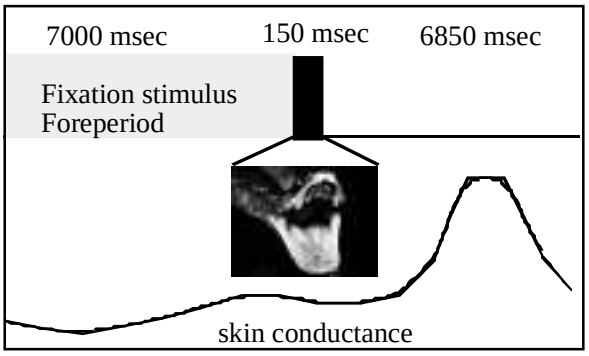


Fig. 1. Timing of stimuli in Animal Fear Study

For the remaining 48 subjects the presentation time of the stimulus was 6 seconds, but in order to keep the total measuring time similar, the fore-period was reduced to 2 seconds. The data of these subjects could not be used for the analysis of the baseline effects because the fore-period was too short to investigate differences. (However, we will consider these data when explaining the importance of having a longer fore-period when establishing true baselines in the results section.) Skin conductance was sampled with a sampling rate of 10 Hz and a resolution of 0.001 microSiemens.

3.3 Data reduction

[1] The description of this experiment closely follows the description in the original publication.

To prevent potential bias from selective analysis of a specific period before stimulus onset, we deliberately defined the "baseline conductance" as the average of the conductance over the **whole** fore-period of 7 seconds. Thus all 70 samples before stimulus onset were reduced to a single value (by averaging) as opposed to using a optimal period based upon visual inspection of the graphs. For each subject, we compared these average values for the erotic pictures with the values for the calm pictures, and then calculated a z-score using the random permutation method (Blair & Kaminski, 1993)[2]. This was also done for the difference between the baseline preceding the snake/spider stimuli and the calm stimuli.

3.4 Data analysis

The hypothesis was tested by comparing the z-scores obtained from each subject with the expected mean z-score of 0 using a simple one-sample t-test.

3.5 Results

Figure 2 shows skin-conductance as a function of sample number in a superposed epoch analysis graph for all subjects combined.

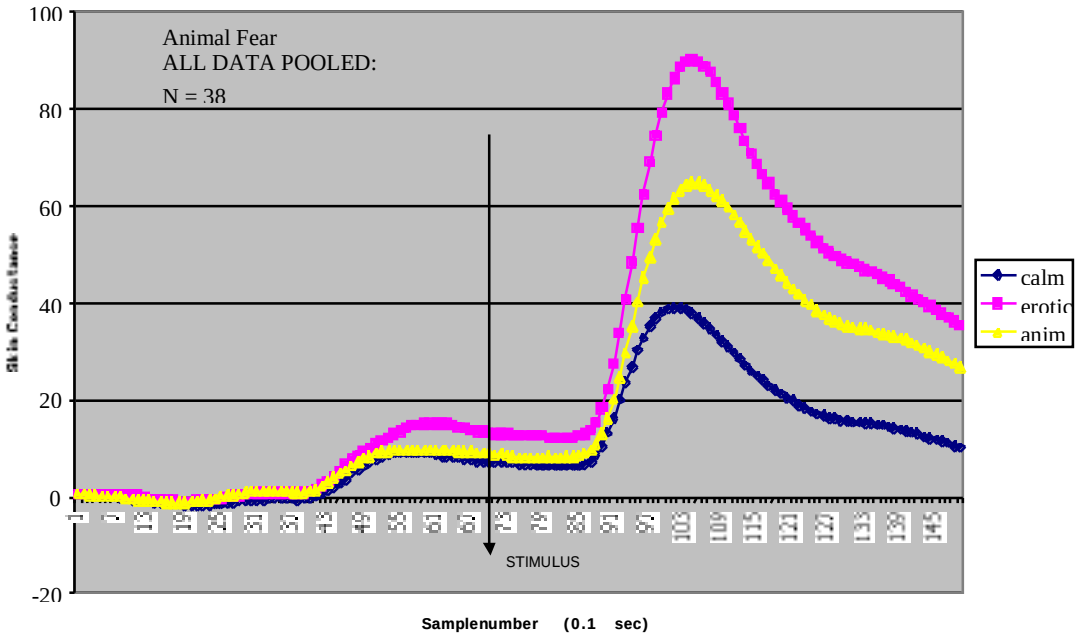


Fig. 2. Average skin conductance for the three type of stimuli.

[2] The RPA method, also known as bootstrap analysis, makes no assumptions about the distribution or the interdependence of the data because it uses the empirical distribution derived from the actual data of each subject. Two procedures may be used to calculate a z-score using this RPA method. In the first method, the empirical mean and standard deviation obtained by repeated simulation of the experiment are used. In the second method, the z-score is calculated from the p-value that arises through counting the number of times the same or a higher score arises in the simulations. We used the latter method because the former method assumes normality. The two methods applied to our data yield z-scores that correlate with a correlation coefficient of ~ 0.99 and thus will produce over-all results which are almost identical.

As can be seen from Fig. 2 it appears as though skin conductance preceding the erotic pictures was larger than skin conductance preceding the snake/spider pictures as well as the level preceding the calm pictures. The skin conductance data were clamped at "-7" seconds (7 seconds before stimulus onset or sample number 1).

3.5.1 erotic versus calm baselines

Out of the 38 subjects, 3 did not show any response at all. They were removed from the statistical analysis. Table 1 shows the calm vs. erotic results. To illustrate the possible impact of person-variables on the anomalous baseline differences, the results are split for male and female.

Table 1

	Mean z-score	df	t	p[3]
Female	0.320	21	1.903	0.0354
Male	0.062	12	0.215	0.4168
Total	0.225	34	1.497	0.0718

3.5.2 Animals (spider/snake) versus calm baselines

Table 2 gives the calm versus spider/snake results split for male and female.

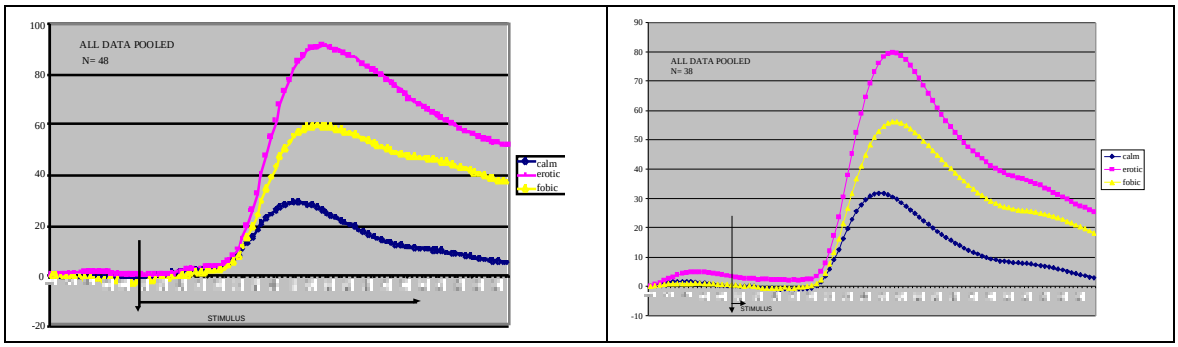
Table 2

	Mean z-score	df	t	p
Female	-0.020	21	-0.135	0.553
Male	0.188	12	0.779	0.226
Total	0.058	34	0.455	0.326

3.5.3 Clamping at -2 seconds

To illustrate the relevance of taking a baseline value at the right moment, and moreover that it is quite difficult to assess anomalous baseline differences if the sampling starts immediately before stimulus onset, we examined the data from subjects that got 6 second exposures in the Animal Fear study (left pane). For these subjects the sampling started 2 seconds before stimulus-onset. We present this plot together with the plot given in Figure 2, but now starting and clamped at 2 seconds before stimulus onset (right pane). Although a slightly larger baseline difference is seen in the 150 msec exposure graph, both graphs look quite similar during the 2 seconds before the stimulus, and it is only because we were fortunate that sampling started much earlier before stimulus onset in the 150 msec exposure that the anomalous baseline differences could be detected.

[3] All p-values are one-tailed because a direction of the expected effect was specified.



4. The gambling study

4.1 Participants

Seven of the participants (4 male, 3 female) were patients with a normal IQ but with bilateral damage to the ventromedial prefrontal cortices. Twelve 5 male and 7 female) were normal control subjects.

4.2 Procedure[4]

In a gambling task simulating real-life decision-making in the way it factors uncertainty, rewards, and penalties, the players (participants) are given four decks of cards, a loan of $2,000 facsimile U.S. bills, and asked to play so that they would lose the least amount of money and win the most. Turning each card carries an immediate reward ($100 in decks A and B and $50 in decks C and D). Unpredictably, however, the turning of some cards also carries a penalty (which is large in decks A and B and small in decks C and D). Playing mostly from the disadvantageous decks (A and B) would lead to an overall loss. Playing from the advantageous decks (C and D) would lead to an overall gain. The players have no way of predicting when a penalty would arise in a given deck, no way to calculate with precision the net gain or loss from each deck, and no knowledge of how many cards they would have to turn to end the game (the game was stopped after 100 card selections). After encountering a few losses, normal participants begin to generate skin conductance responses before selecting a card from the bad decks, and they also begin to avoid the decks with large losses. Patients with stable focal lesions as described above do neither.

4.3 Data extraction and analysis

The data we are interested in are identical to the data used in Bechara's (1997) analyses. This data led to the conclusion that normal participants began to generate anticipatory responses prior to taking a card from one of the disadvantageous decks. From our perspective this anticipatory response is the “baseline” for the response that will follow upon feedback of the actual amount that the participant wins or loses. Rather than evaluating these data as a function of the riskiness of the deck, as was the original goal of the study, we are interested in these baselines as a function of the forthcoming winning or

[4] The description given here is a verbatim copy of the description given by the original authors (Bechara et al, 1997).

losing card. The baseline effect found by Bechara et al can be explained in a normal causal way under the assumption of implicit learning of the riskiness of the decks. However if different baselines are found preceding good or bad cards this would be another example of an anomalous baseline effect. As Bechara et al state explicitly, *"the players have no way of predicting when a penalty will arise…"* (Bechara et al, 1997).

The relevant data of the 12 healthy subjects were extracted (using a graphics pointing device which gives the precise values of the coordinates) from Figure 4 in an article in *Cerebral Cortex* (Bechara et al, 1996), which describes their results. Although this figure gives the values only as a function of sequential order of the card within the deck, we could reconstruct the win/loss amount for each of these cards by using Figure 1 from the article in *Cognition* (Bechara et al, 1994). The data extraction was done by a person blind to the hypothesis. Since the original analyses used parametric tests, we employed a simple t-test to evaluate our hypothesis that baseline levels preceding the losing cards would be higher than those preceding the winning cards.

4.4 Results

Only the data for the normal subjects were used because the brain damaged patients typically did not develop anticipatory responses. Table 3 gives the results of the t-test comparing the baseline skin conductance before winning and before losing cards. The results are also given for each individual deck.

Table 3
Differences between baseline preceding winning and losing cards

Deck	Mean Diff	df	t	P-value
Deck A	0.176	20	1.214	0.1195
Deck B	-0.017	25	-0.102	0.5400
Deck C	0.016	33	0.256	0.3996
Deck D	0.028	33	0.571	0.2860
All Decks	0.085	117	1.634	0.0525

In 3 of the 4 decks the differences are in the expected direction and the decks pooled give a marginally significant difference in the expected direction.

5. The emotional priming study

5.1 Participants

Thirty-two freshman psychology students, 19 female and 13 male, participated in this study.

5.2 Procedure

Each trial consisted of three distinct phases. First, there was a blank screen for 4000 msec, followed by a randomly selected prime which lasted 1000 msec[5]. This was

[5] It should be noted that in half of the trials there was a masked prime with an exposure time of 15 msec. We did not use these data because this prime was not consciously perceived and thus this condition was

followed immediately by a random Japanese character which lasted for 2000 msec after which the subject had to give an evaluation of this character on a 4 point Likert scale (see Figure 3). Skin conductance was sampled with a sampling frequency of 10 Hz and a resolution of 0.005 microSiemens.

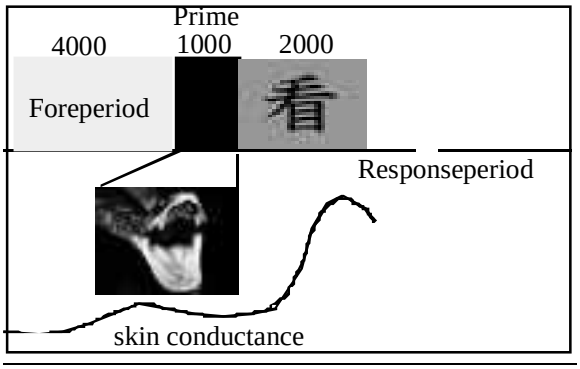


Fig. 3. Timing of stimuli in 'Emotional Priming' Study

5.3 Categories of stimulus material

This study had 5 stimulus categories, not all of which are relevant for the present reanalysis. We removed 2 stimulus categories involving positive and negative faces. These are often used in this type of research, but the emotional content of these is quite moderate compared to the erotic and the fear-inducing stimuli. The three remaining categories more closely match the three categories used in the Animal Fear study, namely, 'erotic', 'threatening animals', and 'neutral' stimuli. The emotional stimuli were each used 6 times and there were 12 neutral stimuli per subject. The presentation order was newly randomized for each subject.

5.4. Data reduction

As in the analysis of the 'Animal Fear study', we used the average of all data preceding the prime as the baseline value. This value was then used to calculate two z-scores per subject in the same way as described before. One z-score represented the mean difference in baseline before the erotic and neutral primes, and the other represented the mean difference in baseline before the threatening animals, primes, and neutral primes.

5.5. Data analysis

A t-test was used to test our main hypothesis that the baselines preceding the emotional primes would be larger than those preceding the neutral primes.

5.6. Results

All data for the 32 subjects and the relevant stimulus categories are shown in Figure 4. The graphic difference between the mean skin conductance records belonging to erotic and neutral primes is quite impressive, but of course they must be tested statistically. It is

---

basically different from the circumstances under which the anomalous baseline effects were originally found.

interesting to note that in this case, as in the Animal Fear study, there is little difference between the records from the fear-inducing primes and the neutral primes. Unlike the Animal Fear study, this also holds for the response part of the records. It should be noted, however, that in the present study no selection of participants was made based on specific animal fear phobias.

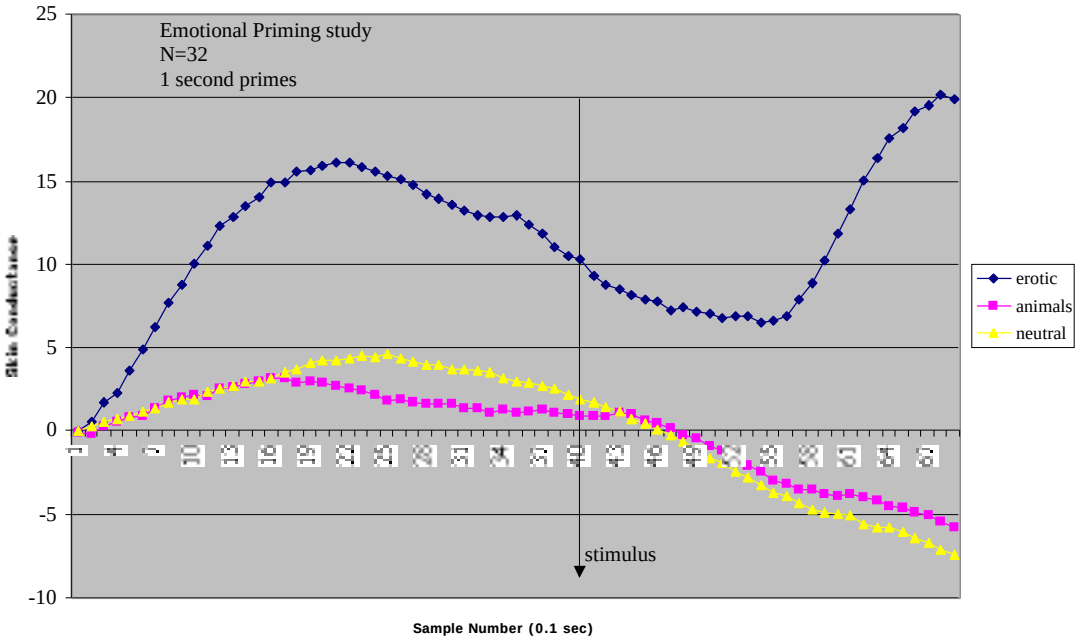


Figure 4: Average skin conductance for the three type of stimuli.

5.6.1. Erotic versus Calm baselines

Table 4 shows the result of the t-test when comparing the mean baseline values for erotic stimuli with those for the neutral stimuli.

Table 4
Results for comparison of baselines preceding erotic and calm pictures.

Stimuli	Mean diff	df	t	p
Erotic - Calm	0.307	31	1.844	0.0374

5.6.2. Evolutionary relevant vs calm baselines

Table 5
Results for comparison baselines preceding 'Animal fear inducing' and calm pictures.

Stimuli	Mean diff	df	t	p
Animal - Calm	0.131	31	0.691	0.2472

6. Conclusions

6.1 Is there an anomalous baseline effect in the data?

In order to answer whether there is an anomalous differential effect that is detectable *prior* to the stimulus presentation, we pooled the results of the three studies using all trials, erotic, animal fear, and calm data. Table VI summarizes all data from the individual experiments, and gives the composite result across the three experiments, calculated by the standard Stouffer Z procedure (Rosenthal, 1978).

Table 6
Overview of 'anomalous' baseline effects in three studies

STUDY	**t (ALL emo – calm)**	**df**	**p**	**Corresponding z**
Animal-fear study	1.44	69	0.0773	1.423
Gambling study	1.634	117	0.0525	1.620
Emotional Priming	1.73	63	0.0431	1.716
OVER-ALL				2.748

Although only one of the studies reached the traditional 5% level for significance, together they certainly seem to support the earlier findings that skin conductance baselines preceding emotional events are higher than those preceding calm events.

The question if this constitutes a genuine anomalous effect in the sense that it cannot be explained within standard scientific models, can be answered only in specifically designed research. All three of these experiments used a randomization method without replacement. This allows for the possibility that subjects might have implicitly learned the ratio between emotional and calm events over the course of the experiment and then they applied this knowledge later in the experiment. Computer simulations modeling this behavior reveal that in the case of randomization without replacement that anticipatory effects as large as a few percent of the main effect can be obtained artifactually (Bierman & Radin, in press).

An argument against this explanation is that such behavior is not observed in practice, especially not the perfectly systematic behavior required to produce slight statistical artifacts. Also, the 'anomalous' effects reported here are much larger than any effect that was obtained in the aforementioned simulations of the effect of randomization without replacement. One further argument could be found in the apparent consistent difference between the baselines preceding emotional stimuli of a different nature, like erotic and animal fear inducing stimuli.

Since it is not the primary goal of this reanalysis to demonstrate the existence of an anomaly, we will not further evaluate this question. Nevertheless, we may conclude that baseline differences prior to a stimulus do occur under generally accepted randomization procedures in psychophysiology. Thus the calculation of response values using such baselines might be in error.

6.2 Assessment of impact of baseline effect on response values[6]

In most psychophysiological research on emotions the variable of interest is the difference in response values after exposure to either emotional or control stimuli. If the response value is calculated as the difference in conductance value at stimulus on-set and the maximum value after some latency time, then the underlying assumption is that the baselines at stimulus onset are not correlated with the emotionality of the upcoming stimulus. As we have shown in the preceding paragraph, this assumption may not be warranted. How serious is this anomalous baseline problem?

In table 7 we have calculated the differences in baseline preceding calm and emotional stimuli at stimulus-onset versus the differences in the peak values after the latency times for the Animal Fear study and the Emotional Priming study. For the Gambling study we have calculated the differences in anticipatory values before taking winning and losing cards versus the differences preceding taking cards from 'good' or 'bad' decks.

Table 7
Comparison of 'anomalous' baseline differences and normal effects

Study	**Baseline difference**	**'Peak value' difference**	**Baselinediff/ 'Peakvalue' diff**
Animal Fear (erotic-calm)	5.87	51.02	11.5%
Gambling (losing-winning)	0.085	0.287	29.6%
Emotional Priming (erotic-calm)	8.33	25.53	32.6%

For the Animal Fear study and the Emotional Priming study this would imply that the response effects of emotionality are *underestimated* by 11.5% and 32.6% respectively if the baselines are assessed at stimulus onset.

The impact of the difference between winning and losing cards on the main effect of interest in the gambling study is more difficult to evaluate. This is because the variable of interest is the anticipatory skin conductance (or what elsewhere would be called the baseline) itself. Here we have an anomalous effect embedded in the “normal” effect. The 29.6% is therefore not an estimate of the error in the calculation of the effect of selecting

[6] It should be noted that the normal anticipatory increase in skin conductance that occurs in paradigms where the subject knows when the stimulus is coming is assumed to be additive to the response when calculating response values by subtracting the baseline at stimulus onset. From the graphs in figure 2 and certainly from the graph in figure 4 we can see that this assumption is suspect. There is a strong indication that the anticipatory contribution declines at stimulus onset and that the contribution to the peak value might be very small. Thus, even without anomalous baseline differences, the assessment of true response values might be contaminated by general anticipation, which influences skin conductance levels at stimulus onset.

"good" versus "bad decks. Detailed examination shows that the 'anomalous' contribution was not evenly distributed over the decks as can be seen in table 8.

Table 8
Anomalous winning-losing card effect split for good and bad decks.

Type of Deck	Mean difference lose-win	df	t-value	p-value
Good	0.043	68	1.246	0.108
Bad	0.160	47	1.675	0.050
total	0.085	117	1.634	0.052

Most of this effect is concentrated in the bad decks (0.16 vs 0.043 for the good decks). Thus the difference between the contribution of the anomalous component is about 0.11, which certainly suggests that the anomalous effect has contributed to some degree to the overall effect.

In conclusion, the analyses provided in this article suggest that whatever the true nature of these apparently anomalous effects, they seem to play a non-trivial role in present day psychophysiological research on emotions.

Acknowledgements

I wish to express my gratitude to A. Hamm and M. Durieux for providing me with the raw data of their experiments. A.Bechara was helpful in pointing out that the randomization used in their study was not appropriate for drawing strong conclusions about the true anomalous nature of the effects. Last, I thank my daughter Jet Bierman for the conscientious way in which she obtained the data from the gambling study.

References

A. Bechara, A. R. Damasio, H. Darnasio, S. W. Anderson, (1994).Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**, pp. 7-15.

A. Bechara, D. Tranel, H. Damasio, A. R. Damasio, (1996).Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb. Cortex* **6**, pp. 215-225.

Bechara, A., Damasio, H., Tranel, D., Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, **275**, 28 February 97, 1293-1295.

Bierman, D. J. & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills*, **84**, 689-690.

Bierman, D. & Radin, D. I. (in press). Anomalous unconscious emotional responses: Evidence for a reversal of the arrow of time. To appear in *Tuscon III: Towards a science of consciousness*, MIT Press.

Blair, R. C. & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, **30**, 518-524.

Boucsein, W. (1992), Electrodermal Activity, pp. 132-161. Plenum Press, New York.

Center for Study of Emotion and Attention (1995*). The international affective picture system: Photographic slides.* Gainesville, Fl.: the Center for Research in Psychophysiology, University of Florida.

Durieux, M. (1999), Master Thesis, University of Amsterdam, dept. Psychology.

Globisch, J., Hamm, A.O., Estevez, F., and Öhman, A. (1999). Fear appears fast: Temporal course of startle reflex potentiation in animal fear subjects. *Psychophysiology*, **36**, pp. 66-75.

LeDoux, J.E. (1996) *The emotional brain.* New York: Simon & Schuster.

Murphy, S.T. & Zajonc, R. B., (1993). Affect cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of personality and social psychology*, **64** (5). pp. 723-739

Radin, D. I. (1997). Unconscious perception of future emotions: An experiment in presentiment. *Journal of Scientific Exploration*, **11** (2), 163-180.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, **85**, 185-193.