Journal of Scientific Exploration, Vol. 11, No. 2, pp. 147-155, 1997 0892-3310/97

The Astrology of Time Twins: A Re-Analysis

CHRISTOPHER C. FRENCH, ANTONY LEADBETTER[1]

Department of Psychology, Goldsmiths' College, University of London, New Cross, London SE14 6NW United Kingdom

GEOFFREY DEAN

Analogic, Box 466, Subiaco 6008, Western Australia

Abstract — Roberts and Greengrass (1994) tested the astrological claim that persons born close together in time are more alike than those born far apart. They collected a total of 128 subjects born on six dates spanning a period of thirty years. Each subject provided their time of birth and completed the short form EPQ. After comparing the interval between births with the difference in EPQ scores the authors concluded that their data showed no strong support for astrology. However, they claimed that the proportion of close personalities increased as the birth interval decreased, which would provide some support for astrology. We have re-analyzed their data and find that their results can be explained as an artifact of data division, sampling errors and age trends. A careful examination of correlations and serial correlations (which are more appropriate tests than the ones used by Roberts and Greengrass) confirmed the absence of astrological effects in their data. If anything, the results were in the wrong direction. A new finding not reported by Roberts and Greengrass was a very small and non-significant difference between younger and older subjects, consistent with the effect of prior knowledge, which suggests that their sample may be slightly contaminated. We conclude that an astrological interpretation of their results is unwarranted.

Keywords: astrology — time twins — post hoc data selection

The Results of Roberts and Greengrass (1994)

In their book, Roberts and Greengrass (1994) tested the astrological claim that "time twins" or persons born close together in time are more alike than those born far apart. This claim is of scientific interest for two reasons. First, it is a logical consequence of the fundamental astrological claim that people resemble their birth charts. Second, it promises to be decisive in astrological research because it avoids the problem of how the various factors in a birth chart should be interpreted and combined. Thus a failure to validate sun signs might be dismissed as a failure to address the correct interpretation or to properly allow for competing factors, but a failure to show that time twins are significantly alike is less easily dismissed. Unfortunately, the evidence for time twins

[1] Now at the Department of Psychology, University of East London

has hitherto been mostly anecdotal and therefore suspect, for example the famous case of Samuel Hemmings and George III is actually fictitious (Dean, 1994). Roberts and Greengrass's study is the first to submit time twins to systematic controlled scrutiny.

Roberts and Greengrass collected, via publicity on the BBC and in national newspapers, a total of 128 subjects born on six dates spanning a period of thirty years (see Table 1). Their mean sample size of 21 is roughly 1% of the 2000 people born every day in the U.K. Each subject provided their birth time and completed the short form Eysenck Personality Questionnaire or EPQ, which measures extraversion (E), psychoticism (P), neuroticism (N), and dissimulation (L), each on a 12-item scale (Eysenck, Eysenck and Barrett, 1985). L is also related to social naivete. For all possible pairs of subjects the authors compared the interval between births with the difference between EPQ scores. According to astrology, the closer the births the closer the personalities and therefore the closer the EPQ scores. The authors (p. 45) concluded that their data showed no strong support for this idea. However, they claimed that the ***proportion*** of close personalities increased as the birth interval decreased, which would provide some support for astrology. This claim was based not on the individual EPQ scores for each pair, but on a single aggregate measure that we will call the EPNL difference, defined as follows:

$$\text{EPNL difference} = \sqrt{(\Delta E^2 + \Delta P^2 + AN^2 + AL^2)}$$

where A is the difference between individual EPQ scores for each pair. For example, their main graph (p. 110) divides the 1400 pairs into four sub-samples and plots the percentage of pairs with close personalities, defined as pairs with an EPNL difference of less than three units. As the mean birth interval increases from 0.5 hour to 24.5 hours the proportion of close personalities decreases from 4.1% to 1.7% (see Table 2). The authors concluded (p. 111) that "the resembling effect is much more powerful for pairs who are born close together than for pairs who merely share the same day," and that this result provides support for astrology.

TABLE 1
Sample Size for Each Date, from Roberts and Greengrass (1994:109)

Birth Date	Age in Years	Sample Size	Pairs
9 Dec 1934	59	16	120
21 Feb 1937	56	21	210
14 Nov 1948	45	33	528
18 Jul 1950	43	17	136
29 Aug 1958	35	23	253
1 May 1964	29	18	153
Total		**128**	**1400**

TABLE 2
Percentage with EPNL Difference < 3 Units vs. Birth Interval*

Mean birth interval (hours)	0.5	3.5	12.5	24.5
% with EPNL differeence < 3 units	4.1	3.1	2.2**	1.7

* From the graphed results of Roberts and Greengrass (1994:110)
**Professor Robersts kindly advised us that this value, plotted on their graph as 2.2, should be 2.5.

Our Re-analysis

Roberts and Greengrass (pp. 105–109) give the birth times and EPQ scores for each of their 128 subjects. Our re-analysis of these data confirmed their conclusion that the difference in EPQ scores was generally unrelated to the birth interval, the mean correlation for the 1400 pairs being –0.001. As a precaution we repeated the analysis excluding 26 subjects (20% of the total) who had L scores of 8 or more, because high L scorers may tend to bias their responses in the direction of social desirability. The results were essentially unchanged.

However our re-analysis did not confirm the authors' second claim that the proportion of close personalities increased as the birth interval decreased. We were able to reproduce the trend shown in Table 2 when we divided the data their way, but this involved very uneven sample sizes for each data point (respectively 98,493,688, and 121 pairs), whereas to equalize the sampling variances we should prefer them to be uniform. In other words, the trend could be an artifact of data division and sampling error. And indeed, when we selected the time intervals and EPNL differences to give uniform sample sizes, the trend was no longer evident (see Table **3**). The results were no better when individual EPQ scores were analyzed, as for E (see Table 4).

Within Days vs. Between Days

Roberts and Greengrass note that the mean EPNL difference measured within the six dates was smaller than the same difference measured between

TABLE 3
EPNL Difference vs. Birth Interval Using Uniform Sample Sizes

Birth Interval (hours)			EPNL Difference							
			0–5.99		6–8.99		9–11.99		> = 12	
Range	Mean	Pairs	N	%	N	%	N	%	N	%
0.0–3.099	1.5	360	104	29	125	35	97	27	34	9
3.1–7.499	5.2	341	104	31	119	35	84	25	34	10
7.5–13.099	10.2	352	102	29	132	38	91	26	27	8
13.1–30	17.8	347	103	30	114	33	99	29	31	9
Does trend support claim?				No		Yes		No		Yes

TABLE 4
E Difference *vs.* Birth Interval Using Uniform Sample Sizes

Birth Interval (hours)			Difference Between Extraversion Scores							
			0–2.99		3–5.99		6–8.99		>=9	
Range	Mean	Pairs	*N*	%	*N*	%	*N*	%	*N*	%
0.0–3.099	1.5	360	123	34	118	33	71	20	48	13
3.1–7.499	5.2	341	117	34	100	29	77	23	47	14
7.5–13.099	10.2	352	127	36	101	29	77	22	47	13
13.1–30	17.8	347	129	37	98	28	78	22	42	12
Does trend support claim?				No		Yes		No		Yes

dates, which they interpret as showing further support for the astrological claim that persons born closer in time are more alike than those born further apart. Their results (p. 111) are as follows:

Mean EPNL difference within dates	7.87	*N* = 1400 pairs
Mean EPNL difference between dates	8.13	N = 1400 pairs drawn at random

When we repeated the same comparison but using all possible pairs between dates, for which N = 6728 pairs, our results confirmed the above observation. The difference within dates was generally smaller than between dates, often very significantly smaller (see Table 5). However, EPQ scores generally change slightly with increasing age (ENP downwards and L upwards), so the difference in scores is bound to increase with increasing age gap, other things being equal. Despite the small sample sizes for each age, such trends occur in the data precisely as predicted (see Table 6, last column), which suggests that Roberts and Greengrass's astrological interpretation of the results is unwarranted.

TABLE 5
Mean Difference Between EPQ Scores Measured Within Dates and Between Dates

	Difference Within Dates		Difference Between Dates			
Measure	Mean	SD	Mean	SD	t	P
E	4.364	3.151	4.505	3.359	1.444	.15
P	1.833	1.405	1.793	1.455	0.941	.35
N	3.828	2.766	4.032	3.029	2.326	.02
L	2.948	2.118	3.209	2.454	3.703	.0002
EN	6.437	3.136	6.688	3.504	2.481	.01
EPNL	7.823	2.929	8.135	3.533	3.091	.002
Pairs	1400		6728		*df* = 8126	

The last two columns indicate the significance (two-tailed) of the difference between means using a t-test.

TABLE 6
Mean EPQ Scores vs. Age of Subjects

Measure	Age in Years of Older Subjects			Age in Years of Younger Subjects			
	59	56	45	43	35	29	*r, df*= 4
E	7.19	5.76	5.67	6.59	7.22	9.33	–.64
P	2.00	1.67	2.18	1.88	2.39	1.78	–.25
N	6.13	4.95	4.73	7.06	6.22	5.28	–.06
L	6.38	5.10	4.85	3.65	4.91	5.17	.42
Sample Size	16	21	33	17	23	18	

The last column shows the correlation between mean EPQ score and age.

Serial Correlation

Roberts and Greengrass relate the supposed astrological connections to planets passing through Gauquelin key sectors (areas of the sky just above the eastern horizon and just past the culminating point), which passage takes on average about two hours per planet. Therefore, as the day progresses, one two-hour period might tend towards E+, the next towards E–, and so on. Two hours is longer than the mean birth interval between adjacent subjects of 1.2 hours for 122 pairs, *i.e.,* excluding pairs involving different dates. For 104 pairs (85%) the birth interval was two hours or less, and for 79 pairs (65%) it was one hour or less, which means that adjacent subjects would tend to share the same planetary connection. Consequently, if the subjects for each date are taken in order of birth, the E score of one should tend to be related to that of the next. In other words, the serial correlation should be positive — a simple test that does not require dividing the sample, but which Roberts and Greengrass neglect to apply. In fact the serial correlations for the four EPQ scores for each of the six dates show no such tendency (see Table 7). In terms of significance level, the two most significant results are $p = .07$ and $p = .08$, which among 4 x 6 = 24 results is the number expected by chance. In terms of direction, there is a very slight preponderance of positive r values over negative r values (13 vs. 11), but the preponderance is reversed as the maximum birth interval is progressively reduced to one hour, which is contrary to what astrology would predict.

For completeness Table 7 also shows the aggregate serial correlation obtained by treating all dates combined as a single sample, excluding of course those pairs involving different dates. The larger sample size increases the sensitivity, but the results are less easily interpreted because the assumption that all dates are from the same population may not hold. However, as they stand, the aggregate correlations are almost exactly at chance level. Their direction tends to agree with that of the mean correlations, otherwise some differences are to be expected due to sampling fluctuations.

As a control, the samples for each date were randomized by redistributing

TABLE 7
Serial Correlations for EPQ Scores When Subjects Are in Birth Order

Age	Sample Size	E	P	N	L	Mean	Pairs	Supportive?* Yes	No	Min p**
59	16	−.216	−.232	−.263	−.477	−.297	15	0	4	.07
56	21	−.165	.270	.243	.200	.137	20	3	1	.25
45	33	−.170	.116	−.257	.043	−.067	32	2	2	.16
43	17	−.263	.322	.128	−.300	−.028	16	2	2	.22
35	23	.161	.032	.132	.076	.100	22	4	0	.47
29	18	.101	−.211	.102	−.441	−.122	17	2	2	.08
Mean		−.092	.050	.014	−.150	−.045	20	13	11	.53
Aggregate		,030	.074	.046	−.014	.034	122			.42
Same but maximum birth interval = 2 hours										
Mean		−.108	.040	.033	−.192	−.057	17	12	12	.46
Aggregate		,011	.083	.043	−.031	.027	104			.40
Same but maximum birth interval = 1 hour										
Mean		−.135	−.090	−.060	−.143	−.107	13	9	15	.64
Aggregate		−.015	−.076	−.053	−.024	−.030	79			.51

*Serial correlation is supportive of astrology if positive.
**Two-tailed p of largest serial correlation (ignoring sign), df = pairs−2.
Aggregate = all subjects combined into a single sample, see text.

the birth times and EPQ scores at random within each date. The mean serial correlation for 10,000 replications was −0.053 with a mean SD of 0.219, in good agreement with the expected values of approximately $-1/(N-1)$ and $1/(N+1)^{1/2}$ respectively, where N = sample size, here 21. When the mean SDs were used to calculate p levels for the means in Table 7, the results were in good agreement with the p levels shown in Table 7.

Contamination by Prior Knowledge of Astrology

Interestingly, when the sample was divided into older and younger subjects, the corresponding Pearson r correlations between the difference in EPQ scores and birth interval were consistently positive and negative respectively, even though the difference was extremely small. When the sample was randomized by redistributing the birth times and EPQ scores at random within each date, the mean Pearson r for 10,000 replications was zero, as expected, and the difference disappeared (see Table 8). None of the observed Pearson r values in Table 8 differ from zero by more than 1.15 of their respective standard deviations (p= .25), so none are even marginally significant. Nevertheless the difference in direction seems too consistent to be easily explained by chance. Indeed, the mean correlations for each of the six dates show a clear tendency to increase with decreasing age despite small sample sizes (see Table 9), the cor-

TABLE 8
Pearson r Correlations Between Difference in EPQ Scores and Birth Times

Subjects		E	P	N	L	EN	EPNL	Mean	p**	Pairs
70 older Subs.	Observed	–.000	–.019	–.058	–.017	–.027	–.029	–.025	*.46*	*858*
(N=16+21+33)	Random*	–.000	.001	.001	–.000	.000	.000	.000		*858*
	SD	.045	.056	.052	.051	.049	.051	.051		
58 younger Subs.	Observed	.047	.019	*,047*	.003	.057	.068	.040	*.35*	*542*
(N= 17+23+18)	Random*	–.000	–.000	–.000	.000	–.000	–.000	–.000		*542*
	SD	.055	.056	.058	.059	.059	.060	.058		

*Pearson r after EPQ scores and birthtimes were redistributed at random within each date, mean of *10,000* replications. Mean SD for *1400* pairs = *0.038.*
**p significance level (two-tailed) of the mean correlation in the previous column, df = pairs –2.

TABLE 9
Pearson r Correlations Between Difference in EPQ Scores and Birth Interval for Each Age Group

	Older Subjects			Younger Subjects		
Age	*59*	*56*	*45*	*43*	*35*	*29*
Mean r	–.090	–.029	–.009	.022	–.026	.177
Pairs	*120*	210	*528*	*136*	*253*	*153*

Each mean r is the mean of the r values for E, P, N, L, EN, and EPNL.

relation between mean r and age being –.79, df= 4, p= .06, almost significant despite the small df.

This result might be explained by prior knowledge of astrology. Given that the subjects were recruited in the name of astrology and were required to know their birth time, and that birth charts and interpretations are readily available, it seems likely that some of them would have known their own birth chart, which knowledge could have biased their EPQ scores in the direction of increased similarity. In this case the results are consistent with a decrease in bias as age increases, as would apply if disinterest in astrology increased with age. This is not implausible, for disinterest in astrology would logically increase as family commitments or other interests take over. Of course the opposite case could also be argued, namely that older persons would have had more time to study astrology and become more knowledgeable, but this would overlook the much lower availability of personal birth charts in their youth. What matters here is not interest in astrology but actual familiarity with their personal birth chart.

In a crude simulation of such prior knowledge effects, we found that the correlation between the difference in say E scores and birth interval was shifted by typically 0.02 when all the E scores in a four-hour window (involving about 20% of the total sample) were increased or decreased by one point. Furthermore, the correlation between E score and sun sign observed in tests of sun-sign prior knowledge is typically around 0.10. Both are compatible with the correlations observed in Table 9, so it seems reasonable to conclude that the sample may be slightly contaminated by prior knowledge.

Analysis by Gauquelin Sector

Roberts and Greengrass claim that the planetary positions in Gauquelin sectors for the top 10% of EPQ scores match the Gauquelin findings. But 10% corresponds to a sample of only 12 subjects, which in terms of statistical power is so weak that it requires a correlation above 0.7 before it can be detected in four out of five tests at a significance level of $p = 0.05$. If a correlation of say 0.8 genuinely existed for 12 subjects, progressively decreasing to 0.0 in the remaining 116, the overall r should have reached the detection limit ($r = 0.25$) of the previous serial correlation tests. But the observed serial r did not reach even 0.05.

Furthermore, of the sixteen possible combinations (4 planets × 4 EPQ scores), Roberts and Greengrass selected their six positive results by inspection, so the role of random fluctuations would seem to be decisive, which point they recognize (p. 112). They attempted to overcome this problem by comparing the pooled distribution for their six positive results with the pooled distribution for the bottom 10% of EPQ scores, obtained by "the same procedure" (p. 112) and therefore presumably also involving selection by inspection, which revealed a difference in accord with the Gauquelin findings. Unfortunately, Roberts and Greengrass do not provide the birthplaces required to calculate accurate sector positions, so we were unable to check their claim.

However, we have seen how the Gauquelin model predicts that, as the day progresses, one two-hour period might tend towards E+, the next towards E–, and so on. We can therefore plot the EPQ scores against birth time to see if the ups and downs coincide with the presence of relevant planets in key sectors. This avoids the problem of selection bias that arises whenever results are selected after the event. We made a provisional attempt using a fixed birthplace centered on the UK, but the results showed no obvious tendency to support the Gauquelin model. This interesting test deserves to be repeated using the actual birthplaces.

Non-EPQ Variables

We were necessarily unable to check variables other than EPQ scores, but the Roberts and Greengrass findings (which are generally negative) are interesting. They gave their subjects further questionnaires, and interviewed and photographed the closest 18 pairs, all born one hour apart or less, to assess similarities. They concluded that there were no clear resemblances in appearance, in handwriting, in names, in interests, in occupation, or in the events in their lives. What similarities existed could be explained as coincidences. However the most compelling similarities existed in the areas of occupation and interests, which they noted was in accord with the Gauquelin findings. The authors conclude that their overall results provide no firm support for traditional astrology but "some support for the new astrology" (p. 78), *i.e.,* the astrology according to Michel Gauquelin. Curiously, the authors' Appendix 2 is devoted to a chart interpretation based on factors previously dismissed as "a litany of failure" (p. 11).

Our Conclusion

Our re-analysis of the data of Roberts and Greengrass found no support for their claim that some astrological effects are discernible in their sample of 128 time twins. Instead their claimed effects can be explained by data division artifacts, misinterpretation of age trends, and misinterpretation of sampling errors, any small residual effects being explained by prior knowledge. Our conclusion is that an astrological interpretation of their results is unwarranted.

References

Dean, G. (1994). Was there ever a Samuel Hemmings? ***Correlation,*** 13, 2, 17.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. ***Personality and Individual Differences,*** 6, 21.

Roberts, P., & Greengrass, H. (1994). The ***Astrology of Time Twins.*** Edinburgh: Pentland Press.

Journal of Scientific Exploration, Vol. 11, No. 2, pp. 157–161, 1997 0892-3310197

Commentary on French *et al.*

PETER ROBERTS

Lark Cottage, Rack Hill, Chalford, Glos GM 8LA, U.K.

French *et al.* have taken much trouble to prove a point which has already been admitted in *The Astrology of Time Twins i.e.* that for the generality of pairs in the data there is no clear correlation between birth time separation and personality resemblance. The possible alternative hypotheses were stated on p. 44 :

1. All those who are born at a particular time share in the 'qualities of that moment of time' and will therefore tend towards a higher resemblance to one another than any of them will towards someone born at a different time.

2. Among those born at a particular time will be a few whose innate characteristics correspond closely to the 'planetary indicators' at that time and place.

The first of these hypotheses was dismissed on p. 45 after showing that the average EPNL separation over all pairs hardly changed as groups with birth time intervals of 0–1, 1–2, 2–3, 3–4 hours were selected. French *et al.* have used an alternative technique (serial correlation) to demonstrate the same point. (They have set up their straw man and then firmly knocked him down).

However, the second hypothesis is of interest — even though it is more difficult to test. It is not a novel hypothesis; it was offered, with a variety of supporting evidence in *The Message of Astrology* published in 1990. (French *et al.* appear to be unaware of it.)

Given that a minority of individuals do have personality traits corresponding to planetary indicators, then it is necessary to examine pairs whose EPNL separation is quite small (referred to in the text as 'close resemblers'). Of all pairs born within an hour, the proportion of close resemblers is substantially higher than the proportion of such close resemblers found among pairs born with a higher birth time separation. Because the number of such pairs involved is quite small, it is necessary to devise the most effective tool practicable. On p. 110 are shown the separate trends for three groups of closely resembling pairs (EPNL<3, between 3 and 4 and between 4 and 5.

The percent of close resemblers is a function of both the birth time separation and of the EPNL distance interval. Thus, multiple regression is appropriate. When the 10 data points shown on p. 110 were subjected to regression analysis, the regression equation found was:

$$z = 1.38 - 0.113x + 1.76y$$

TABLE 1

Predictor	Coefficient	St. Dev.	t-ratio
Constant	1.3758	0.6022	2.28
x	−0.11284	0.02517	−4.48
Y	1.7627	0.1789	9.86

For 7 Degrees of Freedom:

P	t-ratio
1%	3.499
0.1%	5.405

where z = proportion of close resemblers (%)
x = Birth time separation (hours)
y = personality difference (EPNL distance)

To test for significance the t-ratio is needed.

The value of 4.48 for the x coefficient t-ratio lies between that for the p-values of 1% and 0.1%. Hence the comment on p. 111 that "...the statistical significance... is about 300 to 1 against its occurring by chance."

This analysis was not considered appropriate for inclusion in the book, because the book was intended primarily for an essentially lay readership.

French et *al.* state that the finding of a correlation between birth time separation and personality resemblance derives from selection and sampling error. It is true that uneven intervals have been used in the analysis but this was necessitated by the drastic fall-off in numbers of pairs at high birth time separation. (In a 24 hour period, only a small proportion of all birth pairs will be separated by more than, say, 20 hours). Despite the undesirability of uneven intervals, it is very unlikely that the level of significance noted has occurred fortuitously.

The points made by French et *al.* concerning the tests which were attempted to find any association between diurnal positions of planets and personality, are readily admitted. Indeed, the reservations necessary when considering these findings were indicated clearly in the book. However, this was a piece of exploratory research rather than a specific test — and the results, even though only suggestive are of great interest. First, the findings accord quite well with Gauquelin's research on the association of personality traits with diurnal positions of the planets in the natal chart. Secondly, the findings suggest that, given personality tests more closely geared to 'planetary personality profiles', the results of such experiments could well be much more impressive. As a direct result of these time-twin findings, the Astro-Questionnaire Research Group in the U.K. devised experiments based on 'planetary characteristics' (rather than traditional psychological vectors) and showed a very similar result to that found for the time-twin subjects.

The study of time twins carried out for ***The Astrology of Time Twins*** was not regarded as conclusive. In chapter 6 the authors are suitably humble about the importance of their findings, given that it is based on a quite small sample of 128 individuals. It was suggested that for any further progress it will be necessary for the experiment to be repeated, preferably with a larger sample.

Reply to Roberts

CHRISTOPHER C. FRENCH, ANTONY LEADBETTER, AND GEOFFREY DEAN

We thank Professor Roberts for his rejoinder. But it provides no convincing argument against our conclusion, namely that an astrological interpretation of his results is unwarranted. That is, we agree with Roberts that his data contain measurable effects, but we disagree that astrology is needed to explain them. To the specific points raised by Professor Roberts's rejoinder we reply as follows:

1. ***We take much trouble to prove a point already conceded.*** But we devoted only one short paragraph under "Our Re-analysis" to this.

2. ***Hypothesis*** 2 ***says that, of those born at a particular time, a few will have the personality predicted by planetary indicators.*** But this is no different from the null hypothesis, which says hits will happen by chance anyway, so the hypothesis is meaningless unless it predicts ***how many.*** Nevertheless to test it Roberts says "it is necessary to examine pairs whose EPNL separation is quite small." That is, given people born with the same planetary indicators, we cannot examine personality ***vs.*** planets unless we select those who are alike in personality to start with, otherwise we have nothing definite to examine. Fair enough. But the test described by Roberts does not look at personality ***vs.*** planets, it looks at whether the proportion of "alikes" varies with the interval between births.

3. ***The regression equation found was: percent pairs with EPNL difference*** <3 = 1.38–0.113 × ***hours between births*** + 1.76 × EPNL *difference*. This says the proportion of "alikes" decreases with the interval between births, which seems hardly different from hypothesis 1 ("alikeness" decreases with interval), even though hypothesis 1 was ***not confirmed.*** However, if we put EPNL difference = 0 for all pairs, the equation should predict that 100% have an EPNL difference <3, whereas it predicts a maximum of only 1.38%, even falling to ***negative*** quantities for intervals above 1.38/0.113 = 12.2 hours. Similarly awkward is

the prediction that, as the EPNL difference increases, the percent of differences <3 will also increase, when logically it can only decrease. In other words, the equation may fit the data, but in general terms it makes little sense.

4. *The significance of the multiple regression result is* $p = 0.003$, *so it is unlikely to have occurred fortuitously. We* disagree. Look at it this way — each of Roberts and Greengrass's three graphs on their page 110 involves dividing the number of pairs differently, so there are 3 data divisions x 3 EPNL differences = 9 individual regressions that could have been reported, for which our results are presented in Table 1.

Two things stand out. First, look at the number of pairs shown in the top line of figures. Within each graph the number of pairs is divided extremely unevenly (the number per data point varies by a factor of 7, 5 and 14 respectively), which violates the statistical requirement that each data point should have roughly equal variance. Furthermore, as will be evident when the percentages are worked backwards, 12 of the 30 data points involve counts ranging from only 1 to 8, mean 4.7, which cannot help but invite huge sampling errors. The problem is that Roberts's statistical test takes into account only the *number* of data points and ignores their *sample sizes,* whose associated sampling errors could therefore be mistaken for genuine effects. Which is why we used additional tests in our re-analysis.

Second, look at the slopes, especially the asterisked ones. Instead of reporting all 9 regressions, Roberts and Greengrass report *only the best three,* which makes their statistical analysis meaningless. When we submitted all 9 regressions to analysis the resulting equation using Roberts's symbols was $z = 5.3 - 0.042x + 0.23y$. (The z values are percentages, so in principle they should have

Table 1
Regressions from which Best Results were Selected

		Graph 1				Graph 2			Graph 3		
Number of pairs		98	493	688	121	650	629	121	798	544	58
Mean birth interval		0.3	3.2	11.5	21.5	3.0	11.9	21.5	3.9	14.0	23.1
% of pairs	<3	4.1	3.0	2.5	1.7*	3.4	2.2	1.7	3.1	2.2	1.7
with EPNL	3–4	5.1	7.3	6.7	5.8	7.1	6.5	5.8*	6.9	6.2	8.6
differenceof	4–5	4.1	8.5	9.4	6.6	7.7	9.7	6.6	8.6	8.6	5.2*
Slope	<3	−0.10*				−0.09			−0.07		
for EPNL	3–4	−0.00				−0.07*			+0.09		
difference of	4–5	+0.07				−0.06			−0.17*		

* **Indicates the regressions for which Roberts and Greengrass reported results on their page** 110. **Their multiple regression analysis used the midpoints of the birth intervals and EPNL intervals, whereas we used the actual weighted means, *e.g.* for the EPNL intervals they used** 1.5, 3.5 **and** 4.5 **whereas we used** 2.24, 3.44 **and** 4.47. **The significance of our multiple regression wasp** = 0.02, **lower than theirp** = 0.003. **Otherwise the results shown above are much the same as theirs.**

been arcsine transformed to improve normality before being regressed, but in this case it made no appreciable difference.) The factors 0.042 and 0.23 were non-significant (by t test $p = 0.47$ and 0.66, $df = 27$), as was the multiple correlation coefficient R of 0.144, or 0.071 after correction for shrinkage ($p = 0.71$). Finally, in keeping with its nonsignificance, the equation makes little sense, see point (2) above, but less markedly than in Roberts's case. The above results support our conclusion that Roberts's findings are an artifact of data division and sampling error.

6. *Our serial correlation test involves a straw man.* But our test was aimed at the Gauquelin effect (described next), which Roberts says is in accord with his findings. How can this be a straw man?

7. *Gauquelin found an association between personality traits and planets in Gauquelin sectors.* But he found this only for eminent professionals (then about 0.005% of the population), and not for ordinary people. So Roberts's sample of 128 is likely to contain only about $128 \times 0.00005 = 0.006$ such people "at risk." Furthermore the association with personality is now in doubt, for example see Ertel (1993). So what is the more likely explanation for Roberts's findings, a genuine planetary effect or artifacts due to tiny sample sizes and after-the-event selection bias?

8. *Yes, problems due to sampling error and selection bias exist for the Gauquelin sector results.* But Roberts's rejoinder ignores these problems in favor of repeating the very arguments that such problems would invalidate. It also ignores the test that we suggested would avoid selection bias, namely plotting EPQ scores against birth time. Why have arguments when you can have tests?

9. *The experiment should be repeated, preferably with a larger sample.* As it happens one of us has been able to repeat the experiment using cognitive and other variables for a sample of over 5000 adult subjects all born in the same week, and all with birth times recorded to within five minutes or better. The results (which will be reported elsewhere) were uniformly negative.

Reference

Ertel, S. (1993). Why the character trait hypothesis still fails. *Correlation,* **12,** 1, *2.*