RESEARCH

Appraisal of Shawn Carlson's Renowned Astrology Tests

SUITBERT ERTEL

University of Göttingen
Göttingen, Germany
e-mail: sertel@uni-goettingen.de

Abstract—Shawn Carlson's 1985 study, published in Nature, which ended with a devastating verdict of astrology, is scrutinized. The design of Carlson's study violated the demands of fairness and its mode of analysis ignored common norms of statistics. The study's piecemeal analysis of sub-samples avoided testing the totals for astrological effects, as did the neglect of test power, effect size, and sample size. Nevertheless, a correct reanalysis of Carlson's two astrological tests reveals that astrologers matched profiles of the California Personality Inventory to natal charts better than expected by chance with marginal significance (three-way forced choice, p = .054), and that a positive result was replicable by a different assessment method (10-point rating, p = .04). The results are regarded as insufficient to deem astrology as empirically verified, but they are sufficient to regard Carlson's negative verdict on astrology as untenable.

Keywords: astrology—anomalistics—medical flaws—fairness

Introduction and Overview

Among experimental studies of skeptical researchers who aim to test astrological claims, Shawn Carlson's (1985) study, published in *Nature,* the prominent science journal, ended with a devastating verdict on astrology. His study ranked first in subsequent discussions on astrological claims. Today, his paper is referred to on more than 400 Internet pages (Google search), more often than any other paper of its kind. Does Carlson's paper deserve first rank?[1]

Carlson tested two astrological predictions and conducted one study to test the usefulness of his main assessment tool, the California Personality Inventory (CPI).[2] An overview is given in Table 1. Carlson's report is somewhat disorderly; a new numbering of his three studies will be used henceforth.

Study 1

A sample of fewer than 28 astrologers[3] matched 116 birth charts with CPI profiles. CPI profiles are graphs showing 18 scale values, connected by a contour line, on the Y-axis (an interval scale). They represent 18 personality dimensions like dominance, sociability, self acceptance, tolerance, etc., which are aligned on

TABLE 1
Overview of Carlson's Three Studies

Revised division:	Study 1	Study 2	Study 3
Carlson's division	Part 2	Part 1, 2nd section	Part 1, 1st section
Aim of study	Astrological	Astrological	Methodological
Test participants	Astrologers	Students et al.	Students et al.
Invited	N = 90	N = 128	N = ?
Partcipating	N = <28[a]	N = 83	N = 56
	(116)[b]		
Control participants	—	Students et al.	Students et al.
Invited		N = 128	N = ?
Partcipating		N = 94	N = 50
What was to be	CPI profiles[c] with	Personality descrip-	CPI profiles[c] with Self
matched	birth charts	tions[d] with Self	

Note: CPI = California Personality Inventory.
[a] "Some" astrologers initially willing to cooperate refused to continue after checking the task. The number of actually participating astrologers was not given.
[b] In parentheses: Number of CPI profiles that astrologers matched with birth charts.
[c] CPI profiles of students condensing the results of item-wise judgments by 18 scale values.
[d] Personality descriptions obtained from astrologers' interpretations of the students' birth charts.

the X-axis, a categorical scale. The scale values are based on a total of 480 CPI items. The birth charts were divided among the astrologers and three CPI profiles were provided for each chart. One profile was for the chart's owner and the other two were randomly added. The astrologers were asked to match the chart to the CPI profiles and to make first, second, and third choices. They were also asked to rate, on a 10-point scale, the fit of each CPI profile with the birth chart. One astrologer was given a larger number of charts to match them with CPI profiles and to rate their fit. The astrologers had "some background in psychology" and "were familiar with the CPI" (Carlson, 1985: 420). A copy of the CPI interpretation manual (15 pages) was made available in case additional information was needed.

Study 2

Eighty-three volunteer participants, undergraduate or graduate students and others, were each given three different descriptions of personality, which were written by the astrologers who interpreted the persons' birth charts. One of the three descriptions was based on the participants' own birth charts and the other two were added at random from the other chart descriptions in the pool. Each participant was asked how well the descriptions pertained to themselves and to tell which was the first, second, and third best fit. They were also asked to rate the fit of each description on a 10-point scale. The descriptions had been worked out by the astrologers of Study 1. The astrologers' interpretations had been limited to 1000 words and covered the categories of "personality" (300 words), "relationships" (150 words), "career" (150 words), "education" (100 words), and

"current situation" (150 words).[4] Clues related to age, place, time, and astrological concepts were to be avoided. The same material was given to 94 control subjects, matched to the test group with regard to sex, age ("at least three years" difference), and birth date (same sun sign category). Controls were also asked to rank the three personality descriptions with regard to fit to themselves and to rate the fit on a 10-point scale. For the controls, none of the three descriptions was based on their own birth chart.

Study 3

A sample of 56 undergraduate and graduate students and others, who in an earlier session completed the self-administered CPI, were each given three CPI profiles. One of the three profiles was obtained from their own earlier CPI responses, and the other two were randomly selected from the pool of $56 - 1 = 55$ CPI profiles. The participants' task was to choose which profile described their own personality. Unlike participants of Studies 1 and 2, the participants of Study 3 did not also rate the fit on a 10-point scale.

Carlson's Results

Study 1 (astrological aim): According to Carlson, the astrologers ($N < 28$) failed to select, among three CPI profiles, the one that had been obtained from the natal chart owner. He alleged that their judgments on 10-point scales regarding fits of the three profiles to the birth chart did not reveal astrological abilities either. Study 2 (astrological aim): The students ($N = 83$) were not able to select, among the three personality descriptions, the one that was based on their own birth chart, i.e., the astrologer's authentic interpretation was not correctly identified. The selections of personality descriptions by matched controls ($N = 94$) who were presented the same material, of which nothing was related to their own birth charts, were even somewhat better than the test group's selections. Study 3 (methodological aim): The students ($N = 56$) who were presented three CPI profiles, one of them based on their CPI responses, were not able to pick the authentic profile. This was concluded because the controls' ($N = 50$) selections of the same CPI profiles, none of them based on their own CPI responses, did not differ from the test group's selections.

Criticism

Criticism is raised regarding Carlson's data generation (1), his method of data analysis (2), and his neglect of adequate test power (required number of independent observations) (3).

1. Carlson's Procedure of Data Generation

Carlson's personality data were generated by the California Personality Inventory (CPI), occasionally also called the California Psychological Inventory.

Eysenck (1986: 7) objected to Carlson's use of this measurement instrument whose scales he considers "arbitrary and subjective" because they were not chosen by ordinary psychometric techniques. The interpretation of such scales needs, if used at all, "a competent psychologist who has become familiar with this particular device" (quoted from the CPI manual by Hamilton, 1986: 11). None of the participants were psychologists; the experimenter was a graduating physicist. In addition, scale interpretations must take into account the gender of the tested persons, which Carlson did not consider (Eysenck, 1986: 8; Hamilton, 1986: 10). McGrew and McFall (1990: 76) objected, in addition, that "standard psychological tests, like the CPI, may not include the types of information that astrologers require to complete a matching task successfully."

Carlson's test participants were asked to choose one out of three units of personality information (birth chart interpretation in Study 2 or CPI profile in Studies 1 and 3, respectively). A three-choice format, however, is less suited for discrimination than a two-choice format (pair comparison). The advantage of pair comparison is that the complexity of the subject's task is minimized and the precision of the results is increased. Scientists testing a null hypothesis, which they prefer to come true a priori, should provide, following Karl Popper's demand, a fair chance for its refutation. Apparently, Carlson doubted the validity of astrological claims as he developed the protocol of his study. Hence he should have used pair comparison, the fairest existing test format on which Thurstone's "law of comparative judgment" has been based (Thurstone, 1927) and whose methodical aptness is considered as unparalleled in psychophysical research (Gridgeman, 2006). The pair comparison format had been used earlier in another much quoted study by Vernon Clark (1961) that Carlson disregarded.[5]

Another procedural weakness is Carlson's neglect to provide sufficient dissimilarity between the three personality descriptions or CPI profiles, respectively. Supplementing, for multiple choice, the authentic object with two other objects, chosen at random, will generate, by chance, sets of similar objects which are hard to discriminate in the first place. Astrologer Teresa Hamilton, who was willing to cooperate, preferred to resign after having been given three CPI profiles: "I was given some of these charts to match myself, and noticed immediately that the three profiles were often quite similar" (Hamilton, 1986: 12).

An additional flaw in Carlson's protocol is that he did not sufficiently consider the astrologers' methodological suggestions. Astrologer Hamilton, to whom Carlson referred as an advisor giving "valuable suggestions", complained later that "Carlson followed none of my suggestions. I was never satisfied that the experiment was a fair test of astrology" (Hamilton, 1986: 9). Carlson's claim that he followed the astrologers' methodical recommendations is hardly believable since he does not give an account of Hamilton's—and other astrologers'—objections. It is Carlson's misleading account of the role that astrologers had played in his project which earns criticism.

2. *Carlson's Data Analysis*

Carlson's analysis of his three-choice data was also flawed. He tended to analyze the data in a piecemeal fashion, i.e., first, second, and third choice data were treated separately, and total effects were disregarded.

Given the three-choice data, Carlson should have analyzed them, at least, by combining first and second choice frequencies. He himself had intended, "before the data had been analyzed," to combine them: "We had decided to test to see if the astrologers could select the correct CPI profile as either their first or second choice at a higher than expected rate . . ." (p. 425). He ignored his own protocol without giving reasons.

Correct analyses of Carlson's data are required:

Study 1 data: astrologers matching CPI profiles with natal charts. In Study 1, the astrologers matched the subjects' natal charts (N = 116) with CPI profiles. They chose the one that they thought fit best and another one that fit second best from the three supplied profiles. Carlson's Table 2 (Carlson, 1985: 423) provides the following frequencies (Table 2); proportions of authentic choices are also given.

Two charts received a first choice only. The small difference (114 vs. 116) may be neglected. The present analysis is done as if 116 second choices had been made; the number of second choice hits might only have increased (not decreased) if the missing choices had been made.

By combining first and second choice hits, as initially intended by Carlson, we find that 86 hits occurred versus the expected (116) (2/3) = 77.3 hits. The binomial Z-ratio for this result is 1.61, p = .054. Thus, the astrologers chose the correct CPI profile, either as first or second choice, more frequently than expected by chance, at a marginally significant level.

Comment: Carlson does not explain the statistical procedure of his analysis. He uses *standard deviation* as a term to denote both, standard deviations of original distributions and of normalized or Z-distributions, the latter with M = 0 and SD = 1. Readers will be confused by an uncommented use of two different word meanings. For some analyses Carlson seems to calculate confidence intervals of proportions, which differ, however, from the confidence intervals obtained by ordinary procedures. A standard procedure for calculating the confidence interval of proportions dates back to Wilson (1927), of which an account can be found in Newcombe (1998), and in Lowry's VassarStats online (Lowry, 2008).

TABLE 2
Data Obtained from Carlson's Study 1

	Total	Authentic	Proportion authentic
1st choice	116	40	.345
2nd choice	114	46	.404
3rd choice	114	28	.246
1st + 2nd choice	116	86	.741

Using Lowry's procedure, we find that for the expected first + second choice hit proportion (77/116 = .663) the lower and upper limits of .95 confidence are .57 and .75. The observed proportion 86/116 = .74 is near the upper confidence limit (.747) and thus marginally significant. Carlson dealt with the proportion of errors or misses (third choices) instead of first and second choices. The proportion of errors is 28/116 = .240; the limits of .95 confidence for the expected proportion (39/116 = .336) are .253 and .431; the observed proportion (.240) is thus lower than the lower confidence limit (.253) and is therefore significant.

However, the distribution of hits is binomial, not normal; the significance of hits should be obtained from the binomial test (see Lowry, 2008), as shown above. It turns out that in Carlson's Study 1 the binomial Z of hits = 1.61 is significant (p = .05). Carlson incorrectly attributed his Study 1 result to chance, arguing, "Since the rate at which the astrologers chose the correct CPI as their third place choice was consistent with chance, we conclude that the astrologers were unable to chose [sic] the correct CPI as their first or second choices at a significant level" (Carlson, 1985: 425).[6]

An independent analysis based on the astrologers' rating data. In addition to ranking each birth chart into first, second, and third fit categories for their CPI cases, each astrologer "also rated each CPI on a 1–10 scale (10 being highest) as to how closely its description of the subject's personality matched the personality description derived from the natal chart" (Carlson, 1985: 420). Again, Carlson analyzed these ratings piecemeal for each of the three choice categories separately, and found his result "consistent with the scientific prediction of zero slope" (Carlson, 1985: 424). Yet the ratings should have been analyzed together, across choice categories, because they had been made independently of the three ranking choices.[7] Carlson's data, i.e., plot readings from his Figures 3 and 5, are listed in Table 3.[8]

TABLE 3
Frequencies of Astrologers' Ratings 1–10 for 1st, 2nd, and 3rd Choice CPI Profiles: Hits (Authentic Profiles) and Totals (= Authentic Plus Other CPI), Sums for All Choices, and Percentages of Hits

| | 1st choice | | 2nd choice | | 3rd choice | | | | |
| | Carlson Figure 3 | | Carlson Figure 5 | | Carlson Figure 5 | | All choices | | |
Rating	Hits	Total	Hits	Total	Hits	Total	Hits	Total	%
1	0	0	0	0	3	18	3	18	16.6
2	0	0	3	6	7	30	10	36	27.7
3	0	1	2	11	7	27	9	39	23.0
4	2	5	4	13	1	9	7	27	25.9
5	3	7	9	23	4	12	16	42	38.1
6	5	11	12	26	0	1	17	38	44.7
7	7	22	5	14	1	1	13	37	35.1
8	16	43	3	7	0	0	19	50	38.0
9	4	14	0	1	0	0	4	15	26.7
10	1	4	0	0	0	0	1	4	25.0
Sum	38	107	38	101	23	98	99	306	

The percentages from the last column are plotted in Figure 1. A tau-correlation was obtained across N = 306 CPI profiles between the rated fit values (variable X, 1, 2, 3, ... , 10) and the actual hit values (variable Y, 1,0). For example, from 18 ratings of rank 1, 3 were associated with choice 1 (correct choice) and 15 with choice −1 (incorrect choice); from 36 ratings of rank 2, 10 were associated with choice 1 (correct choice) and 26 with −1 (incorrect choice), etc. The nonparametric correlation across 306 paired rating values (1–10) and correctness values (1 and −1) was tau = 0.088, Z = 1.78, p = .037 (one-tailed), ES = .10. The correlation is significant. This result gives reason to take into account the probability that the astrologers were able, to some extent, to successfully match birth charts with CPI profiles.[9] The 10-point rating task (free ratings with allowance for ties) and the three-place matching task (forced choice without ties) are methodically different approaches for testing astrological effects. As such, they are replications, despite changing methods of assessment.

Study 2 data: subjects choosing their own personality descriptions. In Study 2, a test group of subjects (N = 83) tried to choose from three personality descriptions the description that the astrologers had worked out, in Study 1, from their birth charts. They selected from three profiles the one that fitted best and the one that fitted second. The control group (N = 94) was asked to match the same personality descriptions with themselves while none characterized their own personality. Carlson analyzed this data (Table 4) piecemeal, as always,
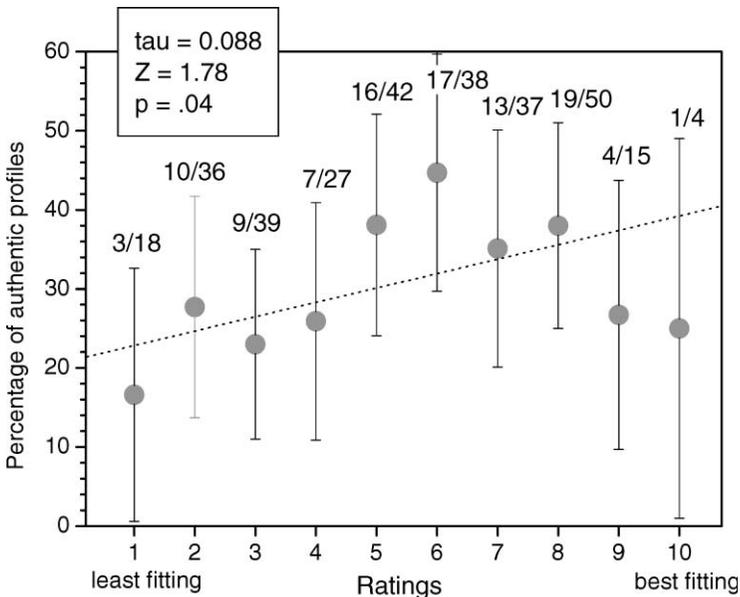

Fig. 1.   Percentage of authentic profiles for 10 points of a rating scale obtained from astrologers who rated three CPI profiles as to how well each matched the natal chart. Limits of 95% confidence. Expected by astrologers, a significant positive correlation; by skeptics, no correlation.

TABLE 4
Numbers of Authentic Personality Descriptions Placed as Best (1), Less Good (2),
and Least Good (3) Fit; Data Obtained from Carlson's Study 2

						Proportion	
	N	1st	2nd	3rd	1st + 2nd	1st + 2nd	3rd
Test group	83	28	33	22	61	.73	.27
Expected					55	.67	.33
Control group	94	42	34	18	76	.81	.19
Expected					63	.67	.33

although first and second choice frequencies should have been combined. One straightforward method of analysis is to compare the difference of correct choice proportions (first + second choice) between the test and control subjects, as follows, in Table 4.

Surprisingly, the direction of the difference is contrary to expectation, yet not significant: $Z = -1.17$, $p = .24$, two-tailed.

Comment: The result of Study 2 is odd. As stated previously, astrologers were successful, to some extent, when they matched CPI profiles with natal charts. Before matching the subjects' profiles, they had interpreted their natal charts (Study 1). In Study 2, the chart owners were asked to choose their authentic descriptions. It is reasonable to expect good quality chart interpretations from astrologers who have successfully matched the same charts with the correct CPI profiles. At least one should not expect the astrologers to produce misleading chart interpretations that would make the students tend to avoid (!) the astrologers' interpretations of their own charts.

Furthermore, personality descriptions with full text are easier to understand than a list of 18 verbal tags with scale values. But, according to Carlson's classification, the test group tended to avoid (!) the astrologers' correct interpretations and instead chose interpretations of the two randomly selected charts so that the control group was even better at this task than the test group, as if they, the controls, had been the genuine subjects.

The result of Study 2 might be due to statistical "chance fluctuation", which is Carlson's claim. A suspicion has been brought forward that Carlson might have mixed up test and control group data, perhaps inadvertently (Joseph Vidmar, personal communication), but this is unverifiable speculation.

Carlson's own analysis of Study 2 data is unacceptable. Instead of calculating the significance of the difference of proportions (first + second choices combined) between test and control groups, he calculated the significance of the deviation of the proportions from expectancy, separately for first, second, and third choices and separately for test and control groups. His procedure ignores the logic of control group designs: Test group frequencies must be compared with control group frequencies. The actual expectancy of the null hypothesis is "no difference between test and control data".

Study 3 data: subjects picking their own CPI profiles. The goal of Study 3 was to find out whether participants who completed the CPI questionnaire (480 items) and who were later presented 18 scale points (profiles) based on their responses to the items, can successfully pick their own profiles from three profiles, two of which, generated by other participants, had randomly been added.

Frequencies and proportions of Study 3 are displayed in Table 5.

The difference between test group and control group proportions has the expected direction, but it is not significant ($Z = 0.589$, p $= .28$).

Comment: By our reanalysis, as shown above, it was found that in Study 1 the astrologers matched CPI profiles to natal charts significantly better than chance. The seeming contradiction between success of Study 1 (astrologers matched CPI profiles to birth charts) and failure of Study 3 (students matched CPI profiles to Self) may be due to unequal amounts of effort which astrologers and students exerted on their tasks. Astrologers were aware of the importance of their test participation; their world views were put on stake. They might have spent much effort on those three CPI profiles of which one belonged to the chart owner while the students might have been less eager to obtain correct results by their matchings and ratings. They might also have been more liable than astrologers to response bias such as social desirability since they had to make judgments about themselves.

3. Test Power of Carlson's Studies

The small number of participants in Carlson's astrological studies is another weakness. For comparison, consider the renowned Gauquelin planetary effects, which showed effect sizes of merely .03 to .07 (Gauquelin, 1988).[10] The effects were generally significant only due to the large sample sizes which amounted to N = 500 to 1000 or more birth data. Even though the effects were small, they nevertheless appeared real because they were statistically very significant. Otherwise, skeptical organizations such as the Comité Para, CSICOP, and CFEPP (Comité Français pour l'Étude des Phénomènes Paranormaux) (Ertel & Irving, 1996), and individuals such as Geoffrey Dean, a strong-minded critic of astrology (Dean, 2000),[11] and H. J. Eysenck, the cautious "metrologist" (Eysenck, 1988),[12] would not have cared about them. Astrological claims should be regarded as being refutable only if the number of independent observations is large enough to reveal

TABLE 5
Numbers of Authentic CPI Profiles Placed as Best (1), Less Good (2),
and Least Good (3) Fit to Self; Data Obtained from Carlson's Study 3

						Proportion	
	N	1st	2nd	3rd	1st + 2nd	1st + 2nd	3rd
Test group	56	25	16	15	41	.73	.27
Control group	50	21	13	16	34	.68	.32

small effects, if such effects exist. The Gauquelin results should be taken as an example for significant astrological correlations.

This is not to say that astrological effects would never exceed the low Gauquelin effect level; our result obtained from Carlson's Study 1 data do not suggest such restriction. The effect size of his result (matching) with p = .054 is ES = .15; the effect size of the complementary result (rating) with p = .037 is ES = .10.[13]

Final Conclusion

The design of Carlson's study was unfair. It was flawed because effect size and test power were not considered. The number of recruited subjects and astrologers was insufficient. Nevertheless, one of Carlson's two astrological studies (Study 1) showed replicable significant results by two methods of assessment.

Studies 2 and 3, for which volunteer students were recruited, seem to have suffered from the participants' lack of effort. Since half of their data (ratings) were so poor that they could not even be analyzed, there is reason to doubt that the analyzable other half of the data has been carefully completed.

In view of at least two significant test results (in Study 1) Carlson's conclusion is not justified: "We are now in a position to argue a surprisingly strong case against natal astrology as practiced by reputable astrologers" (Carlson, 1985: 425). Such conclusion could not be drawn even if Study 1 had shown an *insignificant* result. Not being able to reject a null hypothesis does not justify the claim that the alternate hypothesis is wrong.

As a personal remark: My impression is that research on extraordinary claims, conducted by prominent members of skeptical organizations, when they do such research at all, is likely to suffer from bias and fact-distorting procedures. I noted this repeatedly when I scrutinized studies done under the responsibility of members of the US, Belgian, Dutch, and French skeptical Committees (Ertel, 1995, 1998/1999, 2000, 2005; Ertel & Irving, 1996, 1997).

This is not to say that studies with astrology-favorable results would have an a priori chance to survive my scrutiny any better (Ertel, 1996, 1997, 1998; Ertel & Dean, 1996). Reports about negative results are numerous (Gauquelin, 1955, 1988; Kelly, 1997; Müller & Ertel, 1992; McGrew & McFall, 1990). Confirmations of classical astrological predictions are rare exceptions (Timm & Köberl, 1986; Smit, 1997; Ertel, 2004). Moreover, they do not verify astrological beliefs.[14] Correct astrological diagnoses, if they occur, might be due, e.g., to paranormal intuitions of psi-gifted astrologers (Ertel, 2004). Parapsychological phenomena are likely to eventually become reconcilable with the growing body of scientific knowledge (Radin, 2006); astrology is far from having the same chance. Yet disparaging attitudes against astrology are hardly recommendable as they are articulated in a paper titled "Top Scientists Must Fight Astrology or All of Us Will Face the Consequences" (Crease, 2006). H. J. Eysenck's occasional involvement in astrological research issues may be regarded as a model (Eysenck, 1990; Eysenck & Nias, 1982).

Notes

[1] Shawn Carlson, PhD in physics, is a science writer and chairman of the Society for Amateur Scientists (SAS) which he founded. He is also active as a member of the Committee of the Society of Claims of the Paranormal (CSICOP), and as a supporter of the James Randi Education Foundation (JREF).

[2] The proper name of this test is California Psychological Inventory which is sometimes loosely replaced, as by Carlson, with California Personality Inventory which term will be used here henceforth.

[3] See Note a of Table 1.

[4] Unpublished information taken from Carlson's Guidelines for Preparation of Interpretation.

[5] Vernon Clark's result was astrology-consistent; its evaluation, however, remained controversial (Eysenck, 1984).

[6] For his study, Carlson diverged from statistical conventions by defining significance: "2.5 standard deviation [actually Z] [was] the level we had chosen to call 'significant' ..." (Carlson, 1985: 423). The probability of Z = 2.5 is p = .006, the conventional *first* degree of significance is Z = 1.64, p = .05.

[7] The astrologers had been instructed to do the ratings prior to making the first and second selections: "Compare the chart to each CPI profile. Rate each profile on a scale of 1–10 as to how well each matched the chart overall. Note this on the evaluation sheet. Then select which profile you feel is (1) most likely, (2) second most likely to be the one" (from Carlson's unpublished information sheet).

[8] The numbers based on Carlson's plots of ratings (from his Figures 3 and 5, pp. 421 and 424) differ to a certain extent from his numbers of matchings (in his Table 2, p. 423). His tabulated totals for first, second, and third choice are 116, 114, and 114; his corresponding plotted totals are 107, 101, and 98. His tabulated "correct CPI chosen" are 40, 46, and 28; his corresponding plotted totals are 38, 38, and 23. Carlson does not explain the difference. It seems that some astrologers who completed the matchings did not complete the ratings.

[9] The rather low hit percentages within the two "best fitting" categories might be due to extreme response sets of some astrologers. Because the original data are not available, response sets of the astrologers cannot be taken into account here.

[10] Gauquelin's studies yielded results, of which the "Mars effect" has widely been discussed in public. By numerous replications, Gauquelin obtained significant correlations between birth frequencies of eminent professionals (writers, musicians, painters, physicians, scientists, etc.) and the positions of Mars, Jupiter, Saturn, and Moon at their birth hours. "Effect-sensitive" planetary positions were rise and culmination. Gauquelin's results were confirmed by independently scaling the eminence of professionals (Ertel, 1988) as well as by independent data collection (Müller & Ertel, 1994).

[11] "Gauquelin has covered every possible non-astrological source of error so thoroughly that his results seem beyond doubt. ... [His results] support some of the most fundamental astrological concepts of all ..." (Dean & Mather, 1977: 394).

[12] If the replication should be successful, Gauquelin would rank with the great scientists of the past who changed our conceptions of the Universe; if they should fail, he will be remembered as a brave and honest scientist who was misled by appearance" (Eysenck, 1988: 10).

[13] ES = Z/√N, N = number of independent observations.

[14] Claims of classical astrology (or astrologies) which lack replicable verifications must be sharply distinguished from Gauquelin claims which have been largely verified. The term neo-astrology for Gauquelin's field of studies came into use, which points at this distinction.

Acknowledgments

References

Carlson, S. (1985). A double blind test of astrology. *Nature, 318,* 419–425.

Clark, V. (1961). Experimental astrology. *Aquarian Agent, 1*(9), 22–23.

Crease, R. P. (2006). Top scientists must fight astrology or all of us will face the consequences. *The Scientist*, *3*(5), 9.

Dean, G. (2000). Attribution: A pervasive new artifact in the Gauquelin data. *Astrology under Scrutiny, 13,* 1–72.

Dean, G., & Mather, A. (1977). *Recent Advances in Natal Astrology. A Critical Review 1900–1976.* Bromley & Kent, BRI, UK: Astrological Association.

Ertel, S. (1988). Raising the hurdle for the athletes' Mars effect: Association co-varies with eminence. *Journal of Scientific Exploration, 2,* 53–82.

Ertel, S. (1995). Mars effect survives critique of Dutch skeptics. *Skepter. Skeptische Notities, 8,* 185–203.

Ertel, S. (1996). Re-examination of the gender differences of ordinary people, as claimed by J. F. Ruis. *Correlation, 14*(2), 3–6.

Ertel, S. (1997). Hope for a new science? Scrutiny of Theodor Landscheidt's planetary claims. *Correlation*, *16*(1), 32–39.

Ertel, S. (1998). Scrutiny of Gunter Sachs' excursion into astrological research. *Correlation*, *17*(1), 44–49.

Ertel, S. (1998/1999). Is there no Mars effect? The CFEPP's verdict scrutinized with the assistance of six independent researchers. *Correlation, 17*(2), 4–23.

Ertel, S. (2000). Bulky Mars effect hard to hide. *Journal of Scientific Exploration, 14,* 431–446.

Ertel, S. (2004). Astrologie und Psi. Fallstudie verstärkt die Zusammenhangshypothese. *Zeitschrift für Anomalistik, 4*(1–3), 52–101.

Ertel, S. (2005). Gauquelin planetary effects—Brought down to earth? On Geoffrey Dean's dealing with stubborn facts. *Correlation, 23*(1), 7–33.

Ertel, S., & Dean, G. (1996). Are personality differences between twins predicted by astrology? *Personality and Individual Differences, 21*(3), 449–454.

Ertel, S., & Irving, K. (1996). *The Tenacious Mars Effect.* London: Urania Trust

Ertel, S., & Irving, K. (1997). Biased data selection in Mars effect research. *Journal of Scientific Exploration, 11*(1), 1–18.

Eysenck, H. J. (1984). Methodology and the Vernon Clark experiment. *Astropsychological Problems, 1*(1), 27–29.

Eysenck, H. J. (1986). A critique of "A double-blind test of astrology." *Astropsychological Problems, 4*(1), 7–8.

Eysenck, H. J. (1988). *Foreword to M. Gauquelin: Written in the Stars.* Wellingborough, Northamptonshire, UK: Aquarian Press.

Eysenck, H. J. (1990). *Rebel with a Cause: The Autobiography of Hans Eysenck.* London: Allen.

Eysenck, H. J., & Nias, D. K. B. (1982). A*strology: Science or Superstition?* London: Temple Smith and New York: St. Martin's.

Gauquelin, M. (1955). *L'influence des Astres* [The *Influence* of the Stars]. Paris: Editions du Dauphin.

Gauquelin, M. (1988). *Written in the Stars.* Wellingborough, Northamptonshire, UK: Aquarian Press.

Gridgeman, N. T. (2006). A comparison of some taste test methods. *Journal of Food Science, 26*(2), 171–177.

Hamilton, T. W. (1986). Critique of the Carlson study. *Astropsychological Problems, 3,* 9–12.

Kelly, I. (1997). Modern astrology: A critique. *Psychological Reports, 81,* 1035–1066.

Lowry, R. (2008). VassarStats. Available at: http://faculty.vassar.edu/lowry/VassarStats.html. Accessed March 2009.

McGrew, J., & McFall, R. (1990). A scientific inquiry into the validity of astrology. *Journal of Scientific Exploration, 4,* 75–83.

Müller, A., & Ertel, S. (1992). Astrologisches Zuordnungsexperiment mit Ärzte-Horoskopen [An astrological matching experiment using physicians' horoscopes]. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 34,* 217–221.

Müller, A., & Ertel, S. (1994). *1,083 Members of the French Académie de Médecine.* Waldmohr, Germany: A. P. Müller.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine, 17,* 857–872.

Radin, D. (2006). *Entangled Minds. Extrasensory Experiences in a Quantum Reality.* Simon & Schuster.

Smit, R. (1997). Leo Knegt: A white crow beyond our wildest dreams? *Correlation, 16,* 3–18.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34,* 273–286.

Timm, U., & Köberl, T. (1986). Re-Analyse einer Validitätsuntersuchung an 178 Astrologen. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 28*(1/2), 33–55.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22,* 209–212.