This is a preprint of a paper presented at the 2002 Parapsychological Association Convention in Paris. This document is at http://jeksite.org/others/jw_2002_pa.pdf. The final version is in the "Proceedings of Presented Papers: The Parapsychological Association 45th Annual Convention" pp. 293-305.

# On cumulative effects and averaging artefacts in randomised S-R experimental designs

Jiří Wackermann*

Dept. of Empirical and Analytical Psychophysics,
Institut für Grenzgebiete der Psychologie und Psychohygiene e. V.,
Freiburg i. Br., Germany

## Abstract

Experimental studies into physiological correlates of anomalous cognitive processes (precognitive ESP) or alleged anomalous physiological responses ('presentiment') share a common stimulus-response paradigm. The key components of the experimental strategy are (a) selective averaging of a state variable computed selectively across events of two types ('hits' vs 'misses', 'emotional' vs 'neutral'), and (b) stochastic independence between subsequent events, achieved by randomisation of the stimulus sequences. It is tacitly assumed that, given the null-hypothesis, the mean expectancy of the difference between the conditional averages is zero; 'significant' non-zero differences are then interpreted as indicative of relations between the physiological state and the future event, i. e., of anomalous effects.

This research strategy is untrustworthy and artefact-prone when applied to experimental designs where working memory or expectation effects may play a role. Although the very fact of averaging artefacts due to cumulative effects cannot be denied, practising researchers often tend to ignore the risk of false data-based conclusions, partly due to incomplete understanding of the problem, and partly due to invalid proofs.

In the present paper, the existence of the averaging artefacts is demonstrated on a simple accumulate-and-reset model with a linear accumulation function (so-called 'dinners model'). Combining numerical and analytical approaches, it is shown that (i) the averaging artefact is really present even with 'perfect' randomisation and is not due to inadequate sampling; (ii) the artefact occurs even in 'balanced' experimental scenarios with equal probabilities of events of both types; (iii) the artefact is non-zero for any finite number of stimuli $N$, and vanishes only asymptotically at the rate $N^{-1}$. The analytical approach developed in the paper indicates how detailed analyses of more realistic, complex systems may be carried out.

Averaging artefacts may play a critical role in any experiment involving physiological responses to randomised sequences of stimuli, and can be especially dangerous where the experimental technologies comprise built-in averaging and statistical comparison procedures. Future research should focus on methods to estimate the parameters of the accumulation function in parametric models, on design of statistical procedures to test and separate the pseudo-effects from real effects, and on numerical studies of the properties of such methods. Researchers should resist the temptation of far-reaching conclusions until their data analyses are safeguarded against statistical artefacts.

* Author's address:

Dr. J. Wackermann, Institut für Grenzgebiete der Psychologie und Psychohygiene, Wilhelmstrasse 3a, D-79098 Freiburg i. Br. (Germany). Tel.: +49-761-2072171, fax: +49-761-2072179, e-mail: jw@igpp.de

## 1 Introduction

Experimental designs in 'extra-sensory perception' (ESP) research are mostly based on the stimulus-response (S-R) paradigm. For example, in a 'clairvoyance' ESP experiment, the participant's task may be to identify a hidden target, e. g., one of five Zener symbols displayed behind an opaque shield. In a 'precognition' experiment, the task may be to predict which one of, say, four pictures will show up in a few seconds on the computer display. What makes the tasks so unusual is the fact that the stimulus is not directly perceivable by usual sensory channels, as it is either *concealed* from the subject, or even *delayed* and following the response. In all other respects these are ordinary identification tasks, using a finite number of stimulus classes, and allowing analysis and interpretation in terms of S-R relationships.

The increasing interest in the neural basis of mental processes elicited a search for detectable neurophysiological correlates of those peculiar cognitive (not necessarily *pre*cognitive) processes, usually subsumed under the notion of ESP. The research focus gradually shifted from the search for a 'physiological index of ESP' (Beloff, 1974), i. e., exploration of neural *correlates*, to a search for directly measurable physiological *responses*. The distinction can be best illustrated by comparison of studies aiming at brain electrical correlates of precognition on the one hand (Warren et al., 1992; Don et al., 1995; McDonough et al., 1998; Lehmann et al., 2000), and 'presentiment' studies (Radin, 1997; Bierman & Radin, 1997, 1999) on the other hand.

In both types of experiments, the participants are presented a series of visual stimuli on the computer screen. In the precognition experiment, their motor actions are registered as predictive responses to the (future) stimuli, which may be correct ('hit') or not ('miss'); at the same time, the responses start a pre-programmed mechanism selecting randomly the target and exposing the stimulus on the screen after a constant time delay. In the presentiment experiment, the participants' task is merely to press a button which starts the randomised selection and display of emotionally neutral or highly emotional pictures. The participants are not instructed to predict the type of the picture to be shown; rather, it is their physiology what is expected to respond to the (future) stimulus.

Proper randomisation is important to ensure statistical independence between subsequent stimuli and thus to preclude any conscious or subconscious inference from the preceding stimuli to the next stimulus. This is why researchers paid such intense attention[1] to the implementation and properties of random number generating (RNG) algorithms (Radin, 1985) and to *post hoc* tests of randomness of generated stimuli sequences (Radin, 1997). Long period, minimum serial correlation RNG algorithms to generate the stimulus sequences, with no restrictions on number of stimuli in each class ('open deck') are the current research standard.

Hereinafter we will disregard the differences between particular experimental designs and focus on what they have *in common.* The abstract structure of the experiments is very similar: the subject is exposed to a random series of events which are of binary nature, negative or positive feedback on 'misses' or 'hits' in the precognition task, 'neutral' or 'emotional' pictures in the presentiment experiment. We may call them 'insignificant' and 'significant' events, as these two event types distinctly differ in their psychological or psychophysiological relevance for the subject. However, in order to avoid confusion with the statistical notion of 'significance', we will call them simply *type-0* and *type-1* events, respectively.

[1]As for the intimate relations between the idea of randomised experimental designs and early phases of ESP research, see the highly interesting paper by Hacking (1988).

Another feature 'precognition' and 'presentiment' studies have in common is the statistical data processing strategy: some physiological data (EEG, GSR, etc.) are recorded synchronously with the series of events and selectively averaged across events of type 0 or 1. Significant differences between those conditional averages are then interpreted as indicative of existing physiological correlates of precognitive ESP, or even physiological reactions to visual stimuli. As long as no other explanation can be discovered, such correlations between the present response and the future event are considered 'anomalous'. Thus the two key components of the experimental strategy reviewed above are

(a) selective averaging across type-0 and type-1 events, and comparison of the conditional averages;

(b) stochastic independence between subsequent events, secured by randomised experimental design.

The tacit assumption behind (a) is that, given the null hypothesis, the mean expected value of the difference between conditional averages is zero, at least with perfectly randomised stimulus sequences (b).

In the present paper we will show that this commonly shared research strategy is untrustworthy and the claimed results may be pseudo-effects based on artefacts. It is shown that, under particular circumstances, the mean expectancy of the difference between conditional averages is non-zero, i. e., the assumption underlying (a) is invalid. Consequently, we will argue that the experimental results may be contaminated by statistical artefacts that mimic 'anomalous' effects.[2]

[2]This paper has a relatively long pre-history. The 'dinners model' discussed below was developed by the author in early 1999 as a kind of thought experiment, apt for illustrating possibly false data-based conclusions in statistical inference. Since then, the model was discussed in early 2000 among colleagues interested in the topic, namely, with Prof. D. Lehmann, Dr. R. C. Croft, and, a few months later, with Dr. J. M. Houtkooper. Independently of the author's model, Croft proposed a similar model of working memory effects to explain effects observed in the study by Lehmann et al. (2000). But it was mainly the exchange with Houtkooper, who originally argued *against* the existence of the artefacts described here, that motivated the author to proceed from numerical studies to a rigorous analytical treatment of the problem, given in the Appendix to this paper. In April and May 2000, results of the numerical and analytical studies of the model were described in two non-public papers and circulated within the closed circle of discussants named above.

At that time the author did not consider the papers worth publishing; in his opinion, they were merely a kind of homework exercise, providing a useful starting point for discussion of more general issues. Only later, in March 2001 the author presented the results in an internal meeting at the IGPP, Freiburg, in a talk dealing with the question '*What is an artefact?*'. At the 44th P. A. Convention in New York, August 2001, a brief informal meeting on experimental designs and methodical artefacts in physiological measurements was held with Drs. R. S. Broughton, E. C. May, S. J. P. Spottiswoode, and the author. Following a request by Spottiswoode and May, the author sent them the May 2000 paper. Then it seemed unlikely that, after the world-wide discussions and exchange, the existence of averaging artefacts could be subject to doubts or controversies any longer.

Surprisingly, when J. Westerlund came to Freiburg in March 2002 (to work in a joint project not related to the topic of the present paper), he told the author about his and Prof. J. Dalkvist's studies of expectation effects in the 'presentiment' paradigm. As it was immediately obvious that the model used by Dalkvist & Westerlund was identical to the 'dinners model', the author has felt himself forced to publish his own analyses: not so much to claim his priority but, rather, to give the exact formula for the artefact size and to demonstrate the advantage of the analytic approach.

## 2 Exposition of the problem

It is a common place of experimental psychology that the classical, 'state-less' S-R paradigm is of very limited applicability. Taking into account the internal state (psychological or physiological) of the organism as a co-determinant of the response, the proper model is

$$\text{response} = f(\text{stimulus}, \text{state})$$

where the state at the $n$-th stimulus presentation is a result of the previous history, i. e., of some or all $n-1$ preceding stimuli and responses. In the model, the internal state is updated according to more or less complex rules and may reflect retrospective (working memory) as well as prospective (expectation) effects[3]. This problem is usually believed to be solved by randomised experimental schedules; it is assumed that, summing over all possible stimulus sequences, possible state-dependent effects average out to zero.

Our point, however, is that even perfect randomising procedures do not alleviate the danger of artefacts and erroneous conclusions. To illustrate the point, we will start with a simple cover story.

> Stephen feels a special affection for Phyllis: each Saturday evening he phones her to invite her for a dinner. She is not much impressed by Stephen's person, but she does not want to injure his feelings, so she invented a convenient strategy. Whenever Stephen phones, she rolls fair dice to determine her response. If the die shows '6', she accepts the invitation; in case of any other outcome, from '1' through '5', she finds a socially acceptable excuse to decline.
>
> Stephen also has his secret habits. On Friday he obtains his weekly pocket money from his father, a constant amount of $10. If Phyllis declines to go for a dinner, he saves the money in a shoe box for the next occasion. If she accepts his invitation, he takes the cash from the box and expends all the money for the dinner with Phyllis.
>
> Stephen's father is fairly scared about his son's unfortunate passion, and secretly keeps track of Stephen's cash reserves. On each Saturday the dad notes the state of Stephen's deposit, and then he observes Stephen's going or not going out. In this way, the father obtains a bi-variate data series which, for example, may look like this:

| cash [$] | 10 | 20 | 30 | 10 | 20 | 30 | 40 | 10 | 20 | ... |
|---|---|---|---|---|---|---|---|---|---|---|
| dinner | no | no | yes | no | no | no | yes | no | yes | ... |

> Having collected enough data points, he submits the data to a simple analysis: he calculates the average state of Stephen's funds separately across successful and failed invitations, and finds that the average money sum on accepted invitations was significantly *higher* than on rejected invitations. — Stephen's father is facing an interpretation problem. Given that Steve did not tell Phyllis anything about his money saving habits, something definitely anomalous seems to be happening. Does Phyllis possess an ESP ability? Was she perhaps scanning Stephen's dinners funds? He has a moral problem, too. Should he tell his son what he has found, or not? And, if so, what is he going to tell him?

[3] Of course, state-dependent responses may occur in any 'systems with memory', which may be even simple physical systems, not only conscious subjects or living systems. Think, for instance, of a physicist carrying out repeated measurements of the resistance of a metal wire probe, varying the voltage applied to the probe (stimulus) and measuring the intensity of the electrical current (response); the temperature of the wire probe may play the role of the state variable.

In the story told above we know the mechanism of Stephen's and Phyllis' dates: Phyllis' decisions are driven by random readings of the dice, invitations are accepted with constant probability $\frac{1}{6}$, and there is no causal, normal or anomalous relationship between Stephen's cash reserves and her decision. Thus we know that any effects observed and conclusions drawn by Stephen's father are necessarily based on an artefact. Nevertheless, we are faced with a problem, too: what exactly was wrong with his statistics and conclusions based on it?

The translation of the cover story to the abstract language of data-generating models is straightforward: this is a model for a simple accumulate-and-reset system with a state variable $S$, which is incremented by a constant on each type-0 event and reset by occurrence of a type-1 event (cf. Appendix, eq. (1)). Then, obviously, the 'father's statistics' correspond to the averages between conditional averages of the state variable calculated across type-1 and type-0 events. As long as the nature of the artefact is not fully transparent to us, we may find ourselves in a situation similar to that of Stephen's father with his statistics. The following is a condensed but representative summary of most the frequent statements selected out of comments on the 'dinner funds problem' made by practising researchers[4]:

1. "*There is no such effect, provided the randomisation has been done correctly.*" According to this opinion, the puzzling effect may occur with some selected scenarios, but there is no systematic effect, i. e., the expected mean value of the acceptation/rejection averages is zero when computed over all possible scenarios. There is no problem at all.
2. "*This is an effect of unbalanced experimental design.*" In this opinion, the artefact is caused by unequal probabilities of the two events: invitation accepted with probability $p = \frac{1}{6}$, rejected with probability $q = \frac{5}{6}$. There would be no systematic artefact if both events occurred with equal probabilities $p = q = \frac{1}{2}$.
3. "*Even if there is a systematic artefact, it applies mainly to short experimental runs.*" In this opinion the artefact vanishes rapidly with increasing run length: with reasonably sized experimental series, effects of this kind will not occur, or their magnitude will be negligeable.

In the following section we will examine the behaviour of the 'dinners model' and check all the assumptions listed above against numerical data.

## 3 Examining the 'dinners model'

### 3.1 Numerical studies

The easiest way to check assumption (1) from the preceding section is to submit it to a direct numerical test. To this purpose, a numerical experiment was carried out with $N = 24$ trials, enumerating all $2^N = 16,777,216$ experimental scenarios. For each scenario, the difference $D$

[4]Since early 1999 the author asked more than a dozen colleagues active in psychophysiology, psychology or medical research for their comments on the 'dinner funds problem'. Most interviewed colleagues had achieved a PhD degree or equivalent; all were scientists in good standing in their research areas; and not one of them considered the 'dinner funds problem' as seriously affecting her/his own research. Unfortunately, at that time the author could not foresee the future controversies related to the problem, and did not compile any formal statistics of the replies, so that the exact distribution of the counter-arguments is not known.

between conditional averages was calculated[5] and weighted by the probability of the scenario. Fig. 1A shows the empirical distribution of the statistics $D$; the exact mean value $\langle D \rangle = 1.152386$ is indicated by the arrow. Since this value was obtained by complete enumeration of the sampling space, this single example is sufficient to refute the assumption that selective sampling effects may be responsible for the artefact in question.

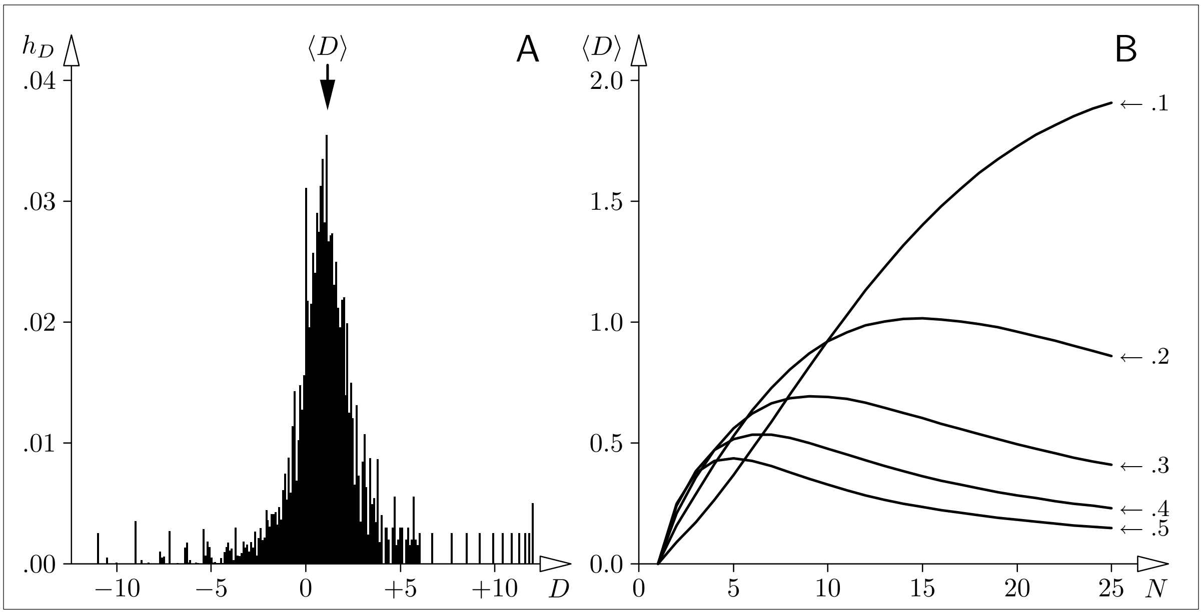


*Figure 1.* A: Distribution of the $D$ statistics for $N = 24$, $p = \frac{1}{6}$; histogram bin width=0.1; mean value $\langle D \rangle$ marked by black arrow. B: Mean expectancies of the $D$ statistics for $1 \leq N \leq 25$, $p$ varied from 0.1 to 0.5; arbitrary units, 1 = state increment on type-0 event.

Exhaustive enumerations over the entire sampling spaces were used also in the next numerical study, where numbers of trials $N$ and probabilities $p$ of the type-1 event were widely and systematically varied; the calculated means $\langle D \rangle$ are plotted in Fig. 1B. The results obviously invalidate the assumption (2) from the preceding section: even with $p = \frac{1}{2}$, the mean value $\langle D \rangle$ deviates from zero. With increasing probabilities $p$ the artifact magnitudes decrease, and their maxima shift to smaller $N$, i. e., shorter experiments; however, the artefact never vanishes.

### 3.2 Analytical solution

The 'brute force' approach applied in the preceding subsection is naturally limited to relatively short experiments. A different approach is required to examine the behaviour of the artefact for large $N$'s. Random sampling (Monte-Carlo simulations) would be the method of first choice. But we saw in section 2 (assumption 1) that the artefact was believed to be a result of *incomplete sampling* of the space of all possible scenarios. Therefore we elaborated an analytic solution, leading to an open-form computational formula for the expectancy of the artefact, $\langle D \rangle$. Details of the mathematical treatment can be found in the Appendix; here we only summarise the qualitative features.

[5]Two singular scenarios, those consisting exclusively of all type-0 or all type-1 events, were inconclusive, as one of the conditional averages was undefined. These special cases were formally assigned $D = 0$, to be consistent with how these scenarios were treated in the analytical part of the paper (Appendix, section 1 and footnote 8).

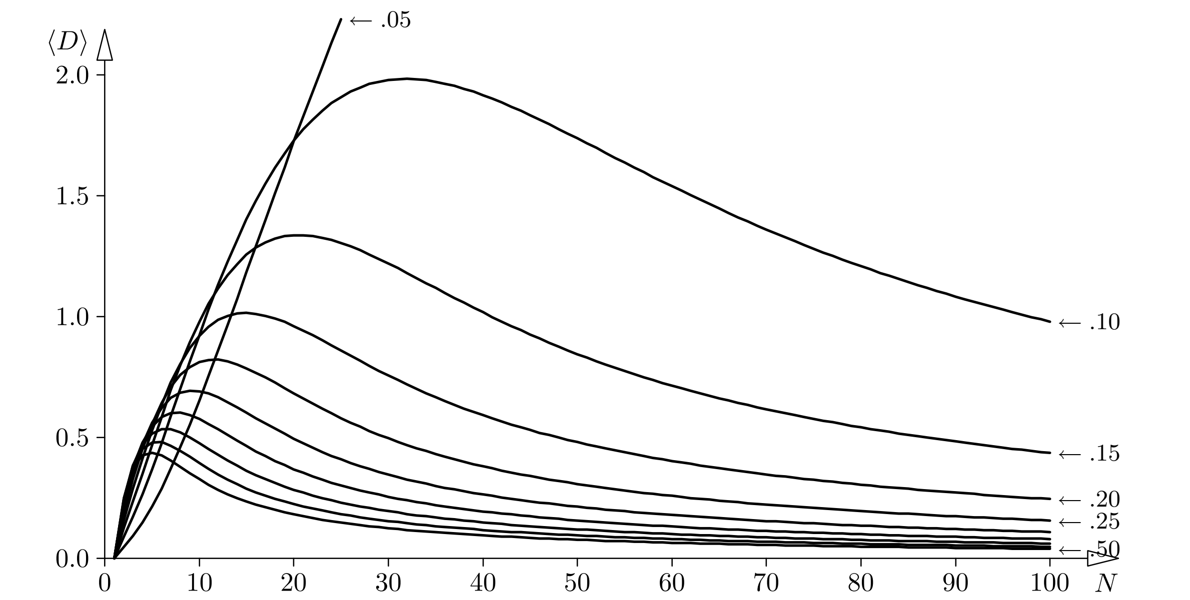


*Figure 2.* Mean expectancies of the $D$ statistics for $1 \leq N \leq 100$, with $p$ varied from 0.05 to 0.50.

Fig. 2 shows the artefact magnitudes as a function of the number of trials $N$ for various values of the type-1 event probability $p$, calculated by the formula (20). Generally, the curves show unimodal courses, with the position and height of the maxima depending on the parameter $p$, as we could see in Fig. 1B; there, however, only short initial segments of the curves could be obtained by the 'brute force' numerical approach. The mean artefact is always positive and vanishes only asymptotically with $N \to \infty$, following an approximate formula

$$\langle D \rangle \approx \frac{1}{Np^2}$$

(hyperbolic 'tails' on the right-hand side).

Thus we see that the averaging artefacts really exist in accumulate-and-reset systems. For the dinners model, representing a special case with linear accumulation function, we have the exact formula[6] for the mean expectancy of the artefact, and also its approximate form for large $N$'s.

## 4 Concluding remarks

Our results invalidate widely spread claims and 'proofs'[7] that sequential artefacts do not exist in perfectly randomised experimental designs. However, we should be very cautious in applying these results to analyses of real experimental data.

[6]An empirically-minded reader could ask whether it was worth the effort to elaborate the exact solution for a model which is an obvious over-simplification. The answer is decidedly affirmative: the mathematical analysis of the simplified model provides guidelines for the treatment of more complex systems of this kind. We follow a well-known dictum of celestial mechanics: "To find an approximate solution of an exact problem, first find the exact solution for an approximation of the problem."

[7]Dr. Bierman provided a formal proof (*http://www.psy.uva.nl/resedu/pn/pubs/bierman/1999/proof.html*), where he claimed that "sequential patterns could not explain anomalous anticipatory physiological behavior", and that considerations of cumulative effects were nothing but a special case of the 'gambler's fallacy'. Regret-

In the analyses presented above, the effect was expressed in arbitrary units, or, more precisely, the increment of the state $S$ on each type-0 event *is* the unit. We cannot tell off-hand the real size of the artefact, measured in physical units; we need procedures to *estimate* the unit of the cumulative effect from the data. Of course, if the averaging artefact were of comparable or greater order of magnitude as the alleged S-R experimental effects, the detectability of the experimental effect would be at stake. The sound scientific attitude then dictates denying claims of any 'anomalous' effects.

In this paper, we have intentionally refrained from in-depth discussions of any particular type of experiment. It is not our task to analyse the consequences of the results for, say, 'presentiment' studies; this is what researchers actively involved in this special research will have to do (Dalkvist, Westerlund, & Bierman, 2002). Although discussions focusing on one particular experimental paradigm may be useful to a limited extent, they cannot serve as a substitute for a thorough treatment of the problem in its general form. On the contrary: general findings resulting from analyses of abstract models should shed light onto special experimental designs and results, and lead to their revision.

However, an in-depth discussion of experimental studies of EEG correlates of precognition or 'presentiment' GSR will also have to take into account the special nature of the response variable, i. e., the electrophysiological data. This is where abstract models of cumulative effects may turn out to be much too crude. For example, in the 'dinners model' the state variable was updated instantaneously, in discrete time moments, and remained static between the updates; in real human physiology experiments, time-varying data (e. g., EEG or EDA curves) are recorded and then averaged over particular time intervals ('sweeps').

Moreover, the state variables in formal models usually have a fixed zero point, which is not necessary the case with the electrophysiology data (floating zero). There we have an additional problem with data pre-processing prior to the averaging: individual data sweeps may be centred on a pre-stimulus average, or clamped to zero at the time of stimulus onset, or at some time preceding the stimulus by a constant interval, etc. All these strategies have their specific pros and cons, and dependences of the effects on these strategies should be studied and understood as well. Obviously, a discussion taking all above-mentioned aspects into account would go far beyond the framework of the present paper, which focused solely on the possibility of the averaging artefacts due to cumulative effects.

Pseudo-effects of this kind should be considered wherever data analyses are based on comparisons between conditional averages, calculated selectively by a binary or multi-valued selector in a randomised experimental schedule. This observation is of special importance for studies using *brain imaging methods*, e. g., the functional magnetic resonance imaging technique

tably, the proof is invalid and the counter-argument misleading. — First, Bierman did not distinguish between relative frequencies and probabilities in his calculations; therefore, his argument applies only to experiments of infinite lengths. Indeed, we have seen that the artefact vanishes *only* with the number of trials $N \to \infty$, but remains non-zero for any finite sequence of stimuli. Second, the 'gambler's fallacy' counter-argument was based on an explicit *category error*. It is true that, in a perfectly randomised experiment, the probability of the next type-1 event is constant, that is, independent of the number of preceding type-0 events and thus also independent of the accumulated internal state, i. e., the gambler's 'subjective expectancy'. From this, however, it does not follow that there is no relation between the type of the next event (0 or 1) and the gambler's internal state at the time of the event. Bierman's counter-argument applies only (and correctly) to the gambler's false belief that he gambler could use readings of his internal state to *predict* the type-1 events.

(fMRI). While in traditional electrophysiology it was the researcher herself/himself who had the data pre-processing and analysis under her/his control and command, the functional brain imaging technologies are usually delivered with ready-made computational procedures for averaging and calculation of between-conditions contrasts; these differential 'effects' are usually visualised by means of statistical probability mapping (SPM) techniques. Although such built-in procedures enhance the researcher's comfort and serve to standardise data analysis, there may be a potential risk of misinterpretation of such statistical contrasts if averaging artefacts are involved.

As mentioned in the Introduction, the discussion of the 'dinners model' was originally aiming at the general issue of erroneous data-based conclusions. The very fact that seemingly trivial systems may display puzzling features and unrecognised artefacts should alarm us. Particularly, we should become more cautious before declaring experimental data indicative of anything 'anomalous'. It is, in the author's opinion, unwise and irresponsible to speculate publicly about revolutionary consequences of findings for our notions of physical reality (cf. Bierman & Radin, 1999), unless all possible sources of artefacts and errors are excluded. Before conclusions are drawn from experimental data, a lot of homework is required to secure the results against artefacts and possible misinterpretations. This statement applies, of course, to any experimental work; but it should be an absolute commandment in highly controversial research like parapsychology. For the sake of our own scientific credit, we should adhere to a simple rule: before we draw any far-reaching conclusions from experimental data, we should be sure that we did our homework timely and correctly.

## 5 Acknowledgements

The author's thanks go to all participants in fruitful discussions on the 'dinners model', and on the nature of sequential artefacts in general, that took place in early 2000: Dr. Rodney C. Croft (then Imperial College School of Medicine, London; now Swinburne University of Technology, Melbourne, Australia), Dr. Joop M. Houtkooper (University of Giessen, Germany), and Prof. Dr. Dietrich Lehmann (University of Zürich, Switzerland). The author also thanks Dr. Robert C. Bishop and Dr. Werner Ehm (IGPP, Freiburg) for their helpful comments on earlier versions of the present paper.

## References

Beloff, J. (1974). ESP: The search for a physiological index. *Journal of the Society for Psychical Research*, **47**, 403–420.

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

Bierman, D. J. & Radin, D. I. (1999). Anomalous unconscious emotional responses: Evidence for a reversal of the arrow of time. In *Toward a Science of Consciousness III* (MIT Press, 1999), pp. 367–386.,

Dalkvist, J., Westerlund, J., & Bierman, D. J. (2002). Simulations of expectation effect in sequential presentations of activating and calm pictures. *Proceedings of the 45th Annual Convention of the Parapsychological Association* (this volume).

Don, N. S., McDonough, B. E., & Warren, C. A. (1995). Event-related brain potential (ERP) indicators of unconscious psi: A replication using subjects unselected for psi. *Journal of Parapsychology* **62**, 127–145.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications.* Vol. I, 3rd ed. New York: J. Wiley.

Graham, R. L., Knuth, D. E., & Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science.* 2nd ed. Reading, Mass.: Addison-Wesley.

Hacking, I. (1988). Telepathy: Origins of randomization in experimental design. *Isis* **79**, 427–451.

Lehmann, D., Pizzagalli, D., Gianotti, L., Faber, P., Wackermann, J., Tanaka, H., Strauch, I., & Brugger, P. (2000). Brain electric correlates of precognition. *International Journal of Psychophysiology* **35**, 42.

McDonough, B. E., Warren, C. A., & Don, N. S. (1998). Third replication of event-related brain potential (ERP) indicators of unconscious psi. *Proceedings of the 41st Annual Convention of the Parapsychological Association*, pp. 64–75.

Radin, D. I. (1985). Pseudorandom number generators in psi research. *Journal of Parapsychology* **49**, 303–328.

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

Warren, C. A., McDonough, B. E., & Don, N. S. (1992). Event-related brain potential changes in a psi task. *Journal of Parapsychology* **56**, 1–30.

## Appendix: The mathematics of the dinners model.

Here we will elaborate an analytic solution for the mean expectancy of the averaging artefact $D$. We will proceed in a more formal fashion than in the article; however, only moderate mathematical skills and basic knowledge of probability theory (e. g., relevant chapters in Feller, 1968) are required to follow the main line of the argument. More on recurrencies and probability generating functions can be found in Graham, Knuth & Patashnik (1994).

### 1 Elements of the formalism

Let $N > 2$ be the number of trials in a S-R experiment. Each trial results in an event $X_n$ ($n = 1, \ldots, N$), which is a binary random variable, $X_n \in \{0, 1\}$. For the sake of abstraction, we will refer to the event $X = 1$ as the 'type-1 event', whatever its interpretation may be (e. g., the positive feedback after a 'hit', exposure of an emotional picture, etc.). In the following we will assume that the probability $p = \mathsf{P}\{X_n = 1\}$ is constant throughout the experiment, and that the events $X_n$ and $X_{n+1}$ are stochastically independent, i. e., the sequence of random variables $X_n$ is a Bernoullian series.

Further let $S$ be a state variable (e. g., physiological) measured in the subject on each trial. Thus the experimental data consists of a series of pairs of observations $(X_n, S_n)$. According to the dinners model, the initial state is $S_1 = 1$, and the state is updated on each trial by the following rules:

$$S_{n+1} = \begin{cases} S_n + 1 & \text{(accumulate if } X_n = 0) \\ 1 & \text{(reset if } X_n = 1) \end{cases} \tag{1}$$

We will call the binary $N$-vectors $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$ experimental *scenarios*. The set of all possible $2^N$ scenarios makes up the complete sampling space $\boldsymbol{\Xi} = \{0, 1\}^N$. Symbols $\mathbf{0}$ and $\mathbf{1}$ denote two singular scenarios consisting entirely of type-0 or type-1 events, respectively. The probability of an individual scenario $\boldsymbol{\xi}$ is

$$P(\boldsymbol{\xi}) = \prod_{\xi_n=0} (1-p) \cdot \prod_{\xi_n=1} p = (1-p)^{N-k} p^k, \tag{2}$$

where $k$ is the number of type-1 events in the scenario $\boldsymbol{\xi}$. The sampling space $\boldsymbol{\Xi}$ can be partitioned into $N + 1$ subsets, each of them consisting of scenarios comprising exactly $k$ type-1 events:

$$\boldsymbol{\Xi} = \boldsymbol{\Gamma}_0 \cup \boldsymbol{\Gamma}_1 \cup \cdots \cup \boldsymbol{\Gamma}_N \tag{3}$$

The numbers of elements of $\boldsymbol{\Gamma}_k$ ($k = 0, 1, \ldots, N$) are given by combinatorial numbers $C_k^N$; as long as the number of trials $N$ is fixed, we will often write simply $C_k$.

Given a scenario $\boldsymbol{\xi}$, we will use shorthand notation for the number of type $j$ events in $\boldsymbol{\xi}$, and the sum of states $S$ across type $j$ trials ($j = 0, 1$):

$${}^{j}N(\boldsymbol{\xi}) = \sum_{\xi_n=j} 1, \quad {}^{j}T(\boldsymbol{\xi}) = \sum_{\xi_n=j} S_n$$

The conditional averages across type $j$ trials are thus

$${}^{j}A(\boldsymbol{\xi}) = \begin{cases} \frac{{}^{j}T(\boldsymbol{\xi})}{{}^{j}N(\boldsymbol{\xi})} & (\text{if } {}^{j}N(\boldsymbol{\xi}) > 0) \\ \text{undefined} & (\text{otherwise}) \end{cases} \tag{4}$$

and the difference between the conditional averages calculated across type-1 and type-0 trials (the 'father's statistics') can be written as[8]

$$D(\boldsymbol{\xi}) = \begin{cases} {}^1A(\boldsymbol{\xi}) - {}^0A(\boldsymbol{\xi}) & \text{(if both terms defined)} \\ 0 & \text{(otherwise)} \end{cases} \tag{5}$$

## 2 Artefact expectancy: General formula

$D$ is a function of elementary scenarios $\boldsymbol{\xi}$ randomly chosen from the sampling space $\boldsymbol{\Xi}$; that is, $D$ is a random variable, distribution of which may depend on the number of trials $N$ and type-1 event probability $p$. Of interest is the expected value

$$\langle D \rangle = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Xi}} P(\boldsymbol{\xi}) D(\boldsymbol{\xi}) \tag{6}$$

For small $N$'s we can evaluate (6) numerically, scanning the entire sampling space $\boldsymbol{\Xi}$, calculating $P(\boldsymbol{\xi})$ and $D(\boldsymbol{\xi})$ for each single scenario. However, this 'brute force' strategy is not applicable with large $N$'s, and a more sophisticated approach is required. Utilising the partition of the sampling space (3), we can rewrite (6) as follows

$$\langle D \rangle = \sum_{k=0}^{N} \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} P(\boldsymbol{\xi}) D(\boldsymbol{\xi}) = \sum_{k=1}^{N-1} \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} P(\boldsymbol{\xi}) \left( \frac{{}^1T(\boldsymbol{\xi})}{{}^1N(\boldsymbol{\xi})} - \frac{{}^0T(\boldsymbol{\xi})}{{}^0N(\boldsymbol{\xi})} \right) \tag{7}$$

For any $\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k$, the index $k$ is equal to the counts of type-1 trials, ${}^1N(\boldsymbol{\xi})$, so we can rewrite (7) into

$$\langle D \rangle = \sum_{k=1}^{N-1} (1-p)^{N-k} p^k \left( \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} \frac{{}^1T(\boldsymbol{\xi})}{k} - \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} \frac{{}^0T(\boldsymbol{\xi})}{N-k} \right) \tag{8}$$

In the following we denote the 'grand totals' over subsets $\boldsymbol{\Gamma}_k$ as

$${}^jG_k = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} {}^jT(\boldsymbol{\xi}) \qquad (j = 0, 1) \tag{9}$$

Substituting this into (8), we have

$$\langle D \rangle = \sum_{k=1}^{N-1} (1-p)^{N-k} p^k \left( \frac{{}^1G_k}{k} - \frac{{}^0G_k}{N-k} \right) \tag{10}$$

In this way we have achieved a separation of terms containing $p$ from terms depending exclusively on $N$ and $k$. Thus, instead of evaluating $P(\boldsymbol{\xi})$ and $D(\boldsymbol{\xi})$ for all $2^N$ scenarios $\boldsymbol{\xi} \in \boldsymbol{\Xi}$, as suggested by (6), we only need to sum over $N-1$ subsets of $\boldsymbol{\Xi}$, provided that we can calculate the terms ${}^jG_k$ easily.

## 3 Calculating the $G$'s

We will start with a search for recursive formulas for the grand totals ${}^jG_k$, defined by (9); then we will define the open-form expressions.

[8] In formula (5) we set $D = 0$ if any of the conditional averages (4) is undefined, i. e., if $\boldsymbol{\xi} = \mathbf{0}$ or $\boldsymbol{\xi} = \mathbf{1}$. This makes sense insofar as the scenarios consisting entirely of 0's or 1's are inconclusive. Another option would be to exclude the offending scenarios $\mathbf{0}, \mathbf{1}$ and to re-normalise the probability measure (2) on the reduced sampling space $\boldsymbol{\Xi} - \{\mathbf{0}, \mathbf{1}\}$; but the convention used here is more natural.

For the shortest possible experiment, $N = 1$, the sampling space consists of merely two scenarios, $\boldsymbol{\Xi} = \{\mathbf{0}, \mathbf{1}\}$; thus, according to (9), we have

$$\begin{array}{ll} {}^0G_0^1 = {}^0T(\mathbf{0}) = 1 & {}^0G_1^1 = {}^0T(\mathbf{1}) = 0 \\ {}^1G_0^1 = {}^1T(\mathbf{0}) = 0 & {}^1G_1^1 = {}^1T(\mathbf{1}) = 1 \end{array} \tag{11}$$

In the following manipulations of recursive formulas, we will need a convenient shorthand notation. In the following we write the length of a scenario, i. e., the number of trials $N$, as the right-hand superscript where appropriate. If $N$ is fixed and we study the $N \to N+1$ transition, then, for the sake of simplicity, we will usually omit the superscript $^N$ and write $^+$ in place of $^{N+1}$. Further we define the *extension operator*, $|$, to add a trial at the end of a scenario $\boldsymbol{\xi} \in \boldsymbol{\Xi}$

$$\boldsymbol{\xi}|\eta = (\xi_1, \ldots, \xi_N)|\eta = (\xi_1, \ldots, \xi_N, \eta) \in \boldsymbol{\Xi}^+$$

Similarly, we define an extension of a subset $A \subseteq \boldsymbol{\Xi}$ by $\eta \in \{0, 1\}$ as

$$A|\eta = \{\boldsymbol{\xi}|\eta : \boldsymbol{\xi} \in A\} \subseteq \boldsymbol{\Xi}^+$$

i. e., the set of all scenarios from $A$ extended by $\eta$. Now, let $\boldsymbol{\Xi}$ and $\boldsymbol{\Xi}^+$ be the spaces of all scenarios of lengths $N$ and $N+1$, respectively. We may write

$$\boldsymbol{\Xi}^+ = \boldsymbol{\Xi} \times \{0, 1\} = \boldsymbol{\Xi}|0 \cup \boldsymbol{\Xi}|1 = \boldsymbol{\Gamma}_0^+ \cup \boldsymbol{\Gamma}_1^+ \cup \cdots \cup \boldsymbol{\Gamma}_N^+ \cup \boldsymbol{\Gamma}_{N+1}^+$$

where

$$\boldsymbol{\Gamma}_k^+ = \begin{cases} \boldsymbol{\Gamma}_0|0 & (\text{for } k = 0) \\ \boldsymbol{\Gamma}_k|0 \cup \boldsymbol{\Gamma}_{k-1}|1 & (\text{for } 1 \le k \le N) \\ \boldsymbol{\Gamma}_N|1 & (\text{for } k = N+1) \end{cases}$$

and thus for $j = 0, 1$ and $1 \le k \le N$ we obtain from (9)

$$^jG_k^+ = \sum_{\boldsymbol{\xi}^+ \in \boldsymbol{\Gamma}_k^+} {}^jT(\boldsymbol{\xi}^+) = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} {}^jT(\boldsymbol{\xi}|0) + \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_{k-1}} {}^jT(\boldsymbol{\xi}|1) \tag{12}$$

To evaluate (12), we need expressions for ${}^jT(\boldsymbol{\xi}|\eta)$. Since the state $S$ is accumulated over consecutive type-0 events and reset by a type-1 event, the state $S_N$ at the end of an experimental scenario $\boldsymbol{\xi}$ is equal to the number of trailing 0's in $\boldsymbol{\xi}$, which we will denote $z(\boldsymbol{\xi})$. Then obviously

$$\begin{array}{ll} {}^0T(\boldsymbol{\xi}|0) = {}^0T(\boldsymbol{\xi}) + z(\boldsymbol{\xi}) + 1, & {}^0T(\boldsymbol{\xi}|1) = {}^0T(\boldsymbol{\xi}) \\ {}^1T(\boldsymbol{\xi}|0) = {}^1T(\boldsymbol{\xi}), & {}^1T(\boldsymbol{\xi}|1) = {}^1T(\boldsymbol{\xi}) + z(\boldsymbol{\xi}) + 1 \end{array} \tag{13}$$

We will also need grand totals of trailing 0's in subsets $\boldsymbol{\Gamma}_k$ $(k = 0, \ldots, N)$:

$$Z_k = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} z(\boldsymbol{\xi})$$

Utilising the decomposition of $\boldsymbol{\Gamma}_k^+$, as we did above in (12), we find the recursive relation

$$Z_k^+ = \sum_{\boldsymbol{\xi}^+ \in \boldsymbol{\Gamma}_k^+} z(\boldsymbol{\xi}^+) = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} z(\boldsymbol{\xi}|0) + \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_{k-1}} z(\boldsymbol{\xi}|1) = \sum_{\boldsymbol{\xi} \in \boldsymbol{\Gamma}_k} (z(\boldsymbol{\xi}) + 1) + 0 = Z_k + C_k$$

from which it follows that

$$Z_k = C_{k+1} \tag{14}$$

Now we can rewrite (12), inserting (13) and (14) where appropriate, to obtain the recursive relations

$$
\begin{aligned}
{}^0G_k^+ &= {}^0G_k + {}^0G_{k-1} + C_{k+1} + C_k = {}^0G_k + {}^0G_{k-1} + C_{k+1}^+ \quad (15)\\
{}^1G_k^+ &= {}^1G_k + {}^1G_{k-1} + C_k + C_{k-1} = {}^1G_k + {}^1G_{k-1} + C_k^+ \quad (16)
\end{aligned}
$$

Obviously, the recursion rule (16) differs from (15) only in the additive term $C_k^+$ instead of $C_{k+1}^+$; we have seen in (11) that the initial values ${}^1G_1^1 = {}^0G_0^1$ are equal; and, obviously, ${}^1G_0^N = 0$ for any $N$. Thus we need only a formula for the grand totals ${}^0G_k^N$, since the following holds

$$
{}^1G_k^N = \begin{cases} 0 & (\text{for } k = 0) \\ {}^0G_{k-1}^N & (\text{for } 1 \le k \le N) \end{cases} \tag{17}
$$

It can be easily proven that

$$
{}^0G_k^N = \begin{cases} (k+1)\, C_{N-k-1}^{N+1} & (\text{for } 0 \le k < N) \\ 0 & (\text{otherwise}) \end{cases} \tag{18}
$$

solves the recursive formula (15).

## 4 Artefact expectancy: Computational formula and large $N$ approximation

Taking into account the identity (17), we can rewrite the equation (10) as

$$\langle D \rangle = \sum_{k=1}^{N-1} (1-p)^{N-k} p^k \left( \frac{{}^0G_{k-1}^N}{k} - \frac{{}^0G_k^N}{N-k} \right) \tag{19}$$

Applying (18) to (19), we find after some elementary algebraic manipulations that

$$\frac{{}^0G_{k-1}^N}{k} - \frac{{}^0G_k^N}{N-k} = \frac{1}{k+2}\, C_{k+1}^{N+1}$$

So we finally obtain a compact form for the mean expected value of the statistics $D$,

$$\langle D \rangle = \sum_{k=1}^{N-1} (1-p)^{N-k} p^k \frac{1}{k+2}\, C_{k+1}^{N+1} \tag{20}$$

which is sufficiently simple and also well suited for numerical computations[9]. Sample curves of the mean expectancy $\langle D \rangle$, plotted as functions of widely varied number of trials $N$, were shown in Fig. 2. Indeed, calculations of $\langle D \rangle$ by (20) and direct computations by (6) yield exactly the same results (Fig. 3A).

Formula (20) shows that, for $0 < p < 1$ and $N > 1$, the expectancy $\langle D \rangle$ is always positive, since all summands on the right hand are positive: the artefact never vanishes with finite $N$. However, we can find a simple quantitative characterisation of the asymptotic behaviour of $\langle D \rangle$ with $N \to \infty$.

[9] This is not the case with formula (19), in spite of its simple appearance. There, the contributions to the total sum are differences of two fractions, ${}^0G_{k-1}^N/k$ and ${}^0G_k^N/(N-k)$, which exceed by many orders of magnitude the difference itself. Thus, as $N$ increases, errors due to the limited precision in the representation of real numbers play a considerable role.

Consider the number $K = {}^1N(\boldsymbol{\xi})$ of type-1 events in randomly chosen scenarios $\boldsymbol{\xi} \in \boldsymbol{\Xi}$, which is a random variable with binomial distribution

$$\mathsf{P}\{K = k\} = \sum_{k=0}^{N} C_k^N (1-p)^{N-k} p^k \tag{21}$$

In the following we will also write

$$Q = \frac{1}{(K+1)(K+2)} \tag{22}$$

Furthermore let

$$\epsilon = (1-p)^N \frac{N+1}{2} + p^N \frac{1}{N+2} \tag{23}$$

Then, applying the identity

$$C_{k+1}^{N+1} = \frac{N+1}{k+1} C_k^N$$

to (20) and adding (23), we have

$$\langle D \rangle + \epsilon = \sum_{k=0}^{N} C_k^N (1-p)^{N-k} p^k \cdot \frac{N+1}{(k+1)(k+2)} = (N+1) \cdot \langle Q \rangle$$

According to (23), $\epsilon > 0$, and thus we have the lower and upper bounds for the expectancy $\langle D \rangle$:

$$0 < \langle D \rangle < (N+1) \cdot \langle Q \rangle$$

Finally, we examine the behaviour of $\langle D \rangle$ for large $N$'s. By 'sufficiently large' we mean $N$ such that $(1-p)^N \approx 0$, so that $\epsilon$ vanishes and $\langle D \rangle \approx \langle Q \rangle$. It can be shown[10] that, for large $N$

$$\langle Q \rangle \approx \frac{1}{(N+1)(N+2)p^2}$$

and thus

$$\langle D \rangle \approx \frac{1}{Np^2} \tag{24}$$

We have found that, with $N \to \infty$, the artefact size $\langle D \rangle \to 0$ at the rate $N^{-1}$; this corresponds to the hyperbolic tails of the curves shown in Fig. 2. Fig. 3 shows the curves of $\langle D \rangle$ for a few selected $p$'s in logarithmic coordinates; the asymptotes with slope $-1$ correspond to the approximation (24).

[10] We do not go into details which would extend this appendix beyond tolerable bounds. It should suffice to note that $\langle Q \rangle$ can be easily evaluated by means of the probability generating function of the binomial distribution (21), which is $g(z) = (1-p+pz)^N$. Integrating this function twice by $z$ and neglecting terms containing $(1-p)^N$, we obtain the above approximation.

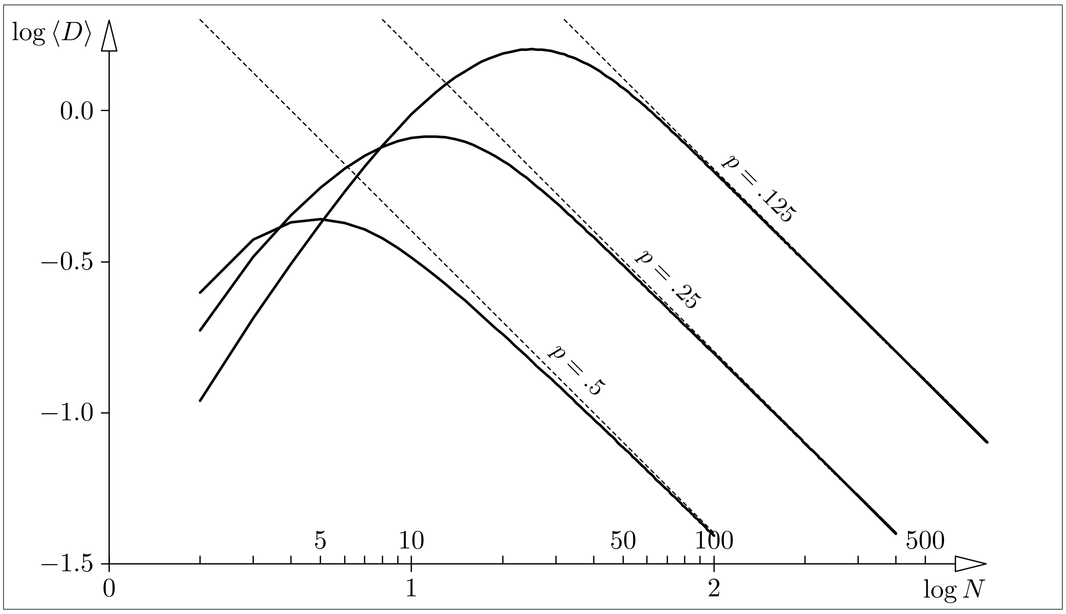


*Figure 3.* Curves of expectancies $\langle D\rangle$ as functions of $N$, for $p = \frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{8}$, plotted in logarithmic coordinates. For large $N$'s, the curves approach the asymptotes $-2\log p - \log N$ (dashed lines).