*Cybernetics And Human Knowing. Vol. 15, no. 2, pp. 49-68*

# Decoupling, Sparsity, Randomization, and Objective Bayesian Inference

*Julio Michael Stern* [1]

Decoupling is a general principle that allows us to separate simple components in a complex system. In statistics, decoupling is often expressed as independence, no association, or zero covariance relations. These relations are sharp statistical hypotheses, that can be tested using the FBST - Full Bayesian Significance Test. Decoupling relations can also be introduced by some techniques of Design of Statistical Experiments, DSEs, like randomization. This article discusses the concepts of decoupling, randomization and sparsely connected statistical models in the epistemological framework of cognitive constructivism.
*Key words:* Bayesian networks, Cognitive constructivism, Covariance structure, Decoupling, Randomization, Sparse factorization.

The light dove, that at her free flight cleaves the air,
therefore feeling its resistance, could perhaps imagine
that she would succeed even better in the empty space.
Immanuel Kant (1724-1804),
in *Critique of Pure Reason* (1787, B-8).

*Step by step the ladder is ascended.*
George Herbert (1593 - 1633),
in *Jacula Prudentium* (1651)

## 1 Introduction

Heinz von Foerster characterizes “known” objects as eigen-solutions for an autopoietic system, that is, as discrete (sharp), separable (decoupled), stable and composable states of the interaction of the system with its environment. Previous articles have presented the Full Bayesian Significance Test (FBST) as a mathematical formalism specifically designed to access the support for sharp statistical hypotheses, and have shown that these hypotheses correspond, from a constructivist perspective, to systemic eigen-solutions in the practice of science, see Stern (2007a). In this article, the role and importance of one of these four essential properties indicated by von Foerster, namely, separation or decoupling, is studied.

Decoupling is the general principle that allows us to understand the world step by step, “looking” at it a piece at a time, localizing single features, isolating basic components or identifying simple objects, out of the immense complexity of the whole universe. In statistical models, decoupling is often introduced by means of no association assumptions, such as independence, zero covariance, etc. In this context,

1. Institute of Mathematics and Statistics, University of São Paulo, Brazil. Email: jstern@ime.usp.br

decoupling relations are sharp statistical hypotheses that can be tested, see for example Stern and Zacks (2002). Decoupling relations in statistical models can also be introduced a priori by means of special Design of Statistical Experiments (DSEs) techniques, the best known of which being randomization.

In Stern (2007b, p. 80) the general meaning of the term *objective* (how, less, more) is defined as the "degree of conformance of an object to the essential properties of an eigen-solution." One of the common uses of the word objective, as opposed to subjective, stresses the decoupling or separation of a given systemic eigen-solution, such as an object of a scientific program, from the peculiarities of a second system, such as a specific human observer. It is this restricted meaning, focusing on the decoupling property of systemic eigen-solutions, that justifies the use of the term objective in this article's title.

The decoupling principle, and one of its most celebrated examples in physics, the vibrating chord, are presented in section 2. In the vibrating chord model, a basic linear algebra operation, the eigen-value factorization, is the key to obtain the decoupling operator. In addition, the importance of eigen-solutions and decoupling operations are discussed from a constructivist epistemological perspective. Herein, we shall focus on decoupling operators related to an other basic linear algebra operation, namely, the Cholesky factorization. In section 3 we show how Cholesky factorization can be used to decouple covariance structure models. In section 4, Simpson's paradox and some strategies for DSEs, such as control and randomization, are discussed. These strategies can be used to induce independence relations, that are expressed into the sparsity structure of the model, which can, in turn, be used for efficient decoupling. In section 5, the role of C.S. Peirce in the introduction of control and randomization in DSEs is reviewed from an historical perspective. This revision will help us set the stage for the discussion, in section 6, of a controversial issue: randomization in Bayesian Statistics. In section 7 some epistemological consequences of randomization, are discussed and the underlying themata of constructivism and objective knowledge are revisited.

The Cholesky factorization operator is presented in section 3, in conjunction with the computational concepts of sparse and structured matrices. Covariance structure and Bayesian networks are some of the most basic and widely used statistical models. Therefore, understanding their decoupling properties is important, not only from a computational point of view, but also from the theoretical and epistemological perspective. Furthermore, one could argue that the usefulness of these statistical models are due exactly to their decoupling properties. Final remarks are presented in section 8.

## 2 The Decoupling Principle

Understanding the entire universe, with all its intricate constituents, relations and interconnections, can be a daunting task, as stated by Schlick:

> The most important (of these) difficulties arises from the recognition of the unending linkage of all natural processes one with another. Its effect is that, on an exact view, every occurrence in the world is dependent on every other; the fall of a leaf is ultimately influenced by the motions of the stars, and

> it would be a task utterly beyond fulfillment to assign its ‘cause’ with absolute completeness to any given process that we suppose determined down to the last detail. For this purpose we should have to adduce nothing less than all of the circumstances of the universe that have so far occurred.
>
> Now fortunately this boundlessness is at once considerably restricted by experience, which teaches us that the reciprocal interdependence of all events in nature is subject to certain easy formulable conditions. (Schlick, 1979, vol.1, p. 292)

Lorenzo Sadun has written an exceptionally clear book on linear algebra, emphasizing the idea of decoupling, that is, the strategy of breaking down complicated multivariate systems into simple *modes*, by a suitable change of coordinates, see also Rijsbergen (2004). Sadun states the goal of his book as follows:

> In this book we cover a variety of linear evolution equations, beginning with the simplest equations in one variable, moving to coupled equations in several variables, and culminating in problems such as wave propagation that involve an infinite number of degrees of freedom. Along the way we develop techniques, such as Fourier analysis, that allow us to decouple the equations into a set of scalar equations that we already know how to solve.
>
> The general strategy is always the same. When faced with coupled equations involving variables $x_1,...,x_n$, we define new variables $y_1,...,y_n$. These variables can always be chosen so that the evolution of $y_1$ depends only on $y_1$ (and not on $y_2,...,y_n$), the evolution of $y_2$ depends only on $y_2$, and so on. To find $x_1(t),...,x_n(t)$ in terms of the initial conditions $x_1(0),...,x_n(0)$, we convert $x(0)$ to $y(0)$, then solve for $y(t)$, then convert to $x(t)$. (Sadun, 2001, p. 1)

As an example of paramount theoretical and historical importance in Physics, we consider the discrete chord. The chord is kept at tension $h$, with $n$ particles of mass $m$ at equally spaced positions $js$, $j$=1...$n$. The extremes of the chord, at positions 0 and $(n+1)s$, are kept fixed, and $\mathrm{x}=[\mathrm{x}_1,\mathrm{x}_2,...,\mathrm{x}_n]'$ denote the particles’ vertical displacements, see French (1974, ch. 5, Coupled oscillators and normal modes, pp. 119-160), Marion (1999, ch. 9) and Franklin (1968, ch. 7), Figure 1 shows the discrete chord for $n$=2.

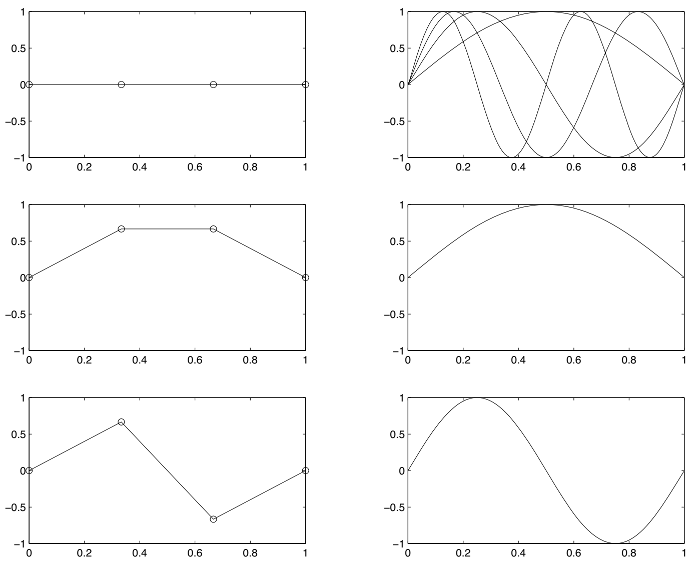


Figure 1: Eigen-Solutions of Continuous and Discrete Chords.

The second order differential equation of classical mechanics, below, provides a linear approximation for the discrete chord system's dynamics:

$$\ddot{x} + Kx = 0, \quad K = w_0^2 \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix}, \quad w_0^2 = \frac{h}{ms}.$$

As it is, the discrete chord differential equation is difficult to solve, since the $n$ coordinates of vector $x$ are coupled by matrix $K$. In the following paragraphs we show how to decouple this differential equation.

Suppose that an orthogonal matrix $Q$ is known to diagonalize matrix $K$, that is, $Q^{-1}=Q'$, and $Q'KQ=D=\text{diag}(d)$, $d=[d_1,d_2,...,d_n]'$. After pre-multiplying the above differential equation by $Q'$, we obtain the matrix equation

$$Q'(Q\ddot{y}) + Q'K(Qy) = I\,\ddot{y} + Dy = 0$$

which is equivalent to the n decoupled scalar equations for harmonic oscillators, $\ddot{y}_k + d_k y_k = 0$, in the new "normal" coordinates, $y = Q'x$. The solution of each harmonic oscillator, as a function of time, t, has the form $y_k(t) = \sin(a_k + w_k t)$, with phase $0 \leq a_k \leq 2\pi$ and angular frequency $w_k = \sqrt{d_k}$ .

The columns of matrix $Q$, the decoupling operator, are the eigenvectors of matrix $K$, which are, as one can easily check, multiples of the un-normalized vectors $z^k$. Their corresponding eigenvalues, $d_k = w_k^2$ , for $j,k=1 \ldots n$, are given by

$$z_j^k = \sin\left(\frac{jk\pi}{n+1}\right), \quad w_k = 2w_0 \sin\left(\frac{k\pi}{2(n+1)}\right).$$

The decoupled modes of oscillation, for $n=2$, are depicted in Figure 1. They are called *normal* modes in physics, *standing* modes in engineering, and eigen-solutions in mathematics. The discrete chord with $n$ particles will have $n$ normal modes, and the limit case, $n \to \infty$ is called the continuous chord. The normal modes of the continuous chord are given by trigonometric functions, the first few of which are depicted in Figure 1. They are also called standing waves or eigen-functions of the chord, and constitute the basis of Fourier analysis.

In either the discrete or the continuous chord, we can *excite*, that is give energy or put in motion, one of the normal modes, without affecting any other normal mode. This is the physical meaning of decoupling, that is, to have separate eigen-solutions. Since the differential equation describing the system is linear, distinct normal modes can also be superposed. This is called the *superposition principle*, which renders the compositionality rule for the eigen-solutions of the chord.

In the original coordinate system, $x$, coupling made it hard to follow the system's evolution. In the normal coordinate system, $y$, based on the system's eigen-solutions, decoupling and superposition made it easier to understand the system behavior. But are these eigen-solutions just a formal basis for an alternative coordinate system, or do they represent "real objects" within the system under study?

Obviously, this is not a mathematical or physical question, but rather an epistemological one. From a constructivist perspective, we can consider these eigen-solutions "objectively known" entities in the system. Nevertheless, the meaning of the term objective in a constructivist epistemology is distinct from its meaning in a dogmatic realist epistemology, as explained in Stern (2006b, 2007a, 2007b).

From a constructivist perspective, systemic eigen-solutions can be identified and named by an observer. Indeed, the eigen-solutions of the vibrating chord have been identified and named thousands of years before mankind knew anything about differential equations. The eigen-values of the chord are known in music as the *fundamental tone* and its *higher harmonics*, and constitute the basis for all known musical systems (see Benade, 1992).

The linear model for the vibrating chord is a paradigmatic example of the fact that, despite the simplicity to understand and manipulate, linear models often give excellent approximations for complex systems. Also, since linear operators are represented by matrices in standard matrix algebra, the importance of certain matrix operations in the decoupling of such models should not be surprising at all. In the vibrating chord model, the eigen-value factorization, $K=QDQ'$, was the key to obtain the decoupling operator, $Q$. The eigen-value factorization plays the same role in many important statistical procedures, such as spectral analysis of time series, wavelet signal analysis, and kernel methods.

Related operations of linear algebra, like Singular Value Decomposition, SVD, and Nonnegative Matrix Factorizations, NNMF, are important in principal components analysis and latent structure models (see e.g., Bertsekas & Tsitsiklis, 1989; Censor & S. A. Zenios, 1998; Cichocki et al., 2006; Dhillon & Sra, 2005; and Hoyer, 2004). Distinct decoupling operators have distinct characteristics, relying upon stronger or weaker structural properties of the model, requiring more or less computational work, and having different capabilities for handling sparse data.

In this article, we will be mainly interested in the decoupling of statistical models. More precisely, we shall focus on decoupling methods related to an important basic linear algebra operation, namely, the Cholesky factorization. In the next section we show how Cholesky factorization can be used to decouple covariance structure statistical models.

The decoupling principle emerges, sometimes with different denominations, in virtually every area of the hard sciences. In systems theory and mathematical programming, for example, it arises under the name of Decomposition Methods. In the optimization of large systems, for example, there are two basic approaches to decomposition:

- High level methods focus on the underlying structure of the optimization problems. High level decomposition strategies replace the original large or complex problem by several hierarchically interconnected small or simple optimization problems, see for example Geoffrion (1972), Lasdon (1970) and Wismer (1971).
- Low level methods look at the matrix representation of the optimization problems. Low level decomposition strategies benefit from tailor made computational linear algebra subroutines to take advantage of the underlying sparse matrix structure. Some of these techniques are discussed in the next section.

## 3 Covariance Structure Models

Covariance structure, multivariate regression, Kalman filter and several other related linear statistical models are widely used in the practice of science. They provide a powerful analytical tool in which the association, coupling or dependence between multiple variables is represented by covariance matrices, as briefly noted in the next paragraphs. These models are simple to manipulate and interpret, and can be implemented using efficient computational algorithms capable of handling millions of (sparsely coupled) variables. In this and the next sections, it is shown how such desirable characteristics of covariance models ultimately rely upon some basic properties of its decoupling operators.

Given a (vector) random variable, $x$, its covariance matrix, $V$, is defined as the expected square distance to its expected (mean) value, $m$, that is,

$$m=E(x),\ V=\mathrm{Cov}(x)=E((x-m)*(x-m)').$$

The diagonal elements, or variances, $Var(x_i)=V_{i,i}$, give the most usual scalar measure of error, dispersion or uncertainty used in statistics, while the off diagonal elements, $Cov(x_i,x_j)=V_{i,j}$ give a measure of association between two scalar random variables, $x_i$ and $x_j$, see Hocking (1985) for a general reference.

Also recall that since the expectation operator, $E$, is linear, that is, $E(Ax+b)=AE(x)+b$ for any random vector $x$, matrix $A$ and vector $b$, we have

$$\mathrm{Cov}(Ax+b)=A\ \mathrm{Cov}(x)A'.$$

The standard deviation, $\sigma_i = \sqrt{V_{i,i}}$, is a dispersion measure given in the same unit as $x$, and the correlation, $C_{i,j} = V_{i,j}/\sigma_i\sigma_j$, is a measure of association normalized in the [-1,1] interval.

As it is usual in the covariance structure literature, we can write a covariance matrix as $V(\gamma) = \sum \gamma_t G^t$, in which the matrices $G^t$ constitute a basis for the space of symmetric matrices of dimension $n$ x $n$, see Lauretto et al. (2002). For example, for dimension $n$=4,

$$V(\gamma) = \sum_{t=1}^{10} \gamma_t G^t = \begin{bmatrix} \gamma_1 & \gamma_5 & \gamma_7 & \gamma_8 \\ \gamma_5 & \gamma_2 & \gamma_9 & \gamma_{10} \\ \gamma_7 & \gamma_9 & \gamma_3 & \gamma_6 \\ \gamma_8 & \gamma_{10} & \gamma_6 & \gamma_4 \end{bmatrix}$$

Using the above notation, we can easily express hypotheses concerning structural properties, including sparsity patterns, in the standard form of vector functional equations, $h(\beta,\gamma)=0$. Details on how to use the FBST to test such general hypotheses in some particular settings can be found in Lauretto et al. (2002).

Once we have established the structural properties of the model, we can estimate the parameters $\beta$ and $\gamma$ accordingly. Following the general line of investigation adopted herein, a question that arises naturally is: How can we decouple the estimated model?

One possible answer to this question can be given in terms of the Cholesky factorization, $LL'=V$ where $L$ is lower triangular. Such a factorization is available for any full rank symmetric matrix $V$, as shown in Golub and van Loan (1989). Let $V=LL'$ be the Cholesky factorization of the covariance matrix, $V$, and let us consider the transformation of variables $y=L^{-1}x$, or $x=Ly$. The covariance matrix of the new variables can be computed as $Cov(y)=L^{-1}VL^{-t}=L^{-1}LL'L^{-t}=I$. Hence, the transformed model has been decoupled, that is, has uncorrelated random components.

Let us consider a simple numerical example of Cholesky factorization:

$$V = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 8 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 2 \end{bmatrix}, \quad V = LL'.$$

This example of Cholesky factorization has some peculiarities: The matrix $V$ is *sparse*, that is, it has several zero elements. In contrast, a matrix with few or no zero elements is said to be *dense*. Matrix $V$ in the example is also *structured*, that is, the zeros are arranged in a nice pattern, in this example, a 2 by 2 off diagonal block. In this example, the Cholesky factor, $L \mid LL'=V$, preserves the sparsity and structure of $V$, that is, no position with a zero in $V$ is *filled* with a non-zero in $L$. A factorization (or elimination) resulting in no fill in is called *perfect*. Perfect eliminations are not always possible, however, there are several techniques that can be used to obtain sparse (and structured) Cholesky factorizations in which the fill in is minimized, that is, the sparsity of the Cholesky factor is maximized. Pertinent references on sparse factorizations include Blair and B. Peyton (1993), Bunch and D. J. Rose (1976) George et al. (1978, 1981, 1989, 1993), Golumbic (1980), Pissanetzky (1984), Rose (1972), Rose and Willoughby (1972), Stern (1992, 1994), Stern and Vavasis (1993, 1994) and van der Vorst and van Dooren (1990).

Large models may have millions of sparsely coupled variables. A sparse and structured factorization of such a model gives a 'simple' decoupling operator, $L$. This is a matter of vital importance when designing efficient computational procedures. In practice, large models can only be computed with the help of these techniques. An other important class of statistical models, Bayesian Networks, relies on sparse factorization techniques that, from an abstract graph theoretical perspective, are almost identical to sparse Cholesky factorization, see for example Lauritzen (2006) and Stern (2006a, sec. 9-11).

In the next section we continue to examine the role of covariance, or more general forms of association, in statistical modeling. On particular, we examine some situations leading to spurious associations, destroying a model's presumed sparsity and structure. In the following sections we review, from an historical and epistemological perspective, some techniques of design of statistical experiments (DSE), used to induce (no) association relations in statistical models. These relations translate into sparsity and structural patterns that, in turn, can be used by efficient factorization algorithms.

## 4 Simpson's Paradox and the Control of Confounding Variables

Lindley (1991, pp. 47-48) illustrates Simpson's paradox with a medical trial example. From 80 patients in the study, 40 received treatment, T, and 40 received a placebo with no effect, NT. Some patients recovered from their illness, R, and some did not, NR. The recovery rates, R%, are given in Table 1, where the experimental data is shown, both in aggregate form for All patients, and separated or disaggregated according to Sex. Looking at the table one concludes that the treatment is bad for either male or female patients, but good for all of them together! This is the Simpson's Paradox: The association between two variables, T and R in Lindley's example, is reversed if the data is aggregated / disaggregated over a *confounding* variable, Sex in Lindley's example.

**Table 1: Simpson's Paradox.**

| Sex | T | R | NR | Tot | R% |
|---|---|---|---|---|---|
| All | T | 20 | 20 | 40 | 50% |
| All | NT | 16 | 24 | 40 | 40% |
| Male | T | 18 | 12 | 30 | 60% |
| Male | NT | 7 | 3 | 10 | 70% |
| Fem | T | 2 | 8 | 10 | 20% |
| Fem | NT | 9 | 21 | 30 | 30% |

Lindley provides the following scenario for the situation illustrated by this example: The physician responsible for the experiment did not trust the treatment and also was aware that the illness under study affects females most severely. Hence, he decided to try it mainly on males, who would probably recover anyway. This illustrates the general Simpson's paradox situation, generated by the association of the confounding variable with both the explained and one (or more) of the explaining variables. Additional references on several aspects related to the Simpson paradox include Blyth (1972), Cobb (1998), Good and Mittal (1987), Gotzsche (2002), Greenland et al. (1999, 2001), Heydtmann (2002), Hinkelmann (1984), Pearl (2004) and Reintjes et al. (2000).

The obvious question then is: How can we design a statistical experiment in order to avoid spurious associations?

Two strategies are self-evident:

- Control possible confounding variables in order to impose some form of invariance (constancy, equality) in the experiment, or

- Measure possible confounding variables so that the relevant ones can be included in the statistical model.

The simplest form of the first strategy would be to test the treatment in a set of *clones*, individuals that are, using the words of Fisher (1966, sec. 9, Randomization; the Physical Basis of Validity of the Test, pp. 17-19), “exactly alike, in every respect except that to be tested.”

This strategy, however, is too strict. Even if feasible, the conclusions of the study would only apply to the clone population, not to individuals from a population with natural variability.

A more general form of the first strategy in known as blocking, defined in Box et al. as:

> The device of pairing observations is a special case of ‘blocking’ that has important applications in many kinds of experiments. A block is a portion of the experimental material (the two shoes of one boy in this example) that is expected to be more homogeneous than the aggregate (all shoes of all the boys). By confining treatment comparisons within such blocks, greater precision can often be obtained. (Box, 1978, pp. 102-103, Sec. 4. 3, Blocking and Randomization)

Blocking is a very important strategy in the design of statistical experiments (DSEs), used to increase, whenever possible, the precision of the study’s conclusions.

As for the second strategy, it looks a sure thing! No statistician would ever refuse more information, in a larger and richer data bank.

Nevertheless, we have to ask whether we want to control and/or measure SOME of the possibly confounding variables, that is those perceived as the most important or even those we are aware of, or ALL of them?

Keeping everything under control in a statistical experiment (or in life in general) constitutes, in the words of Fisher, “a totally impossible requirement in our example, and equally in all other forms of experimentation” (Fisher, 1966, p. 18).

Not only the cost and complexity of trying to do so for a very large set of variables would be prohibitive in any practical circumstance, but also

> it would be impossible to present an exhaustive list of such possible differences (variables) appropriate for any one kind of experiment, because the uncontrolled causes which may influence the result are always strictly innumerable. (Fisher, 1966, p. 55)

Modern theory of DSEs offers a way out of this conundrum that, in its most concise form, see Box et al. (1978, pp. 102-103), can be stated as: Control what you can, and randomize what you can not. Randomization, as defined by Hacking, is

> “(the) notion of random assignment of treatment to a subset of the plots or persons, leaving the rest as controls. ... I shall speak of an experiment using randomization in this way as involving a randomized design. ...
>
> ... There is a related but distinguishable idea of (random) representative sampling.” (Hacking, 1988, p. 428)

As it is usual in the statistical literature, Hacking distinguishes between two intended uses of randomization, namely random design and random sampling. Random design aims to eliminate bias coming from systematic design problems, including several forms of uncontrolled influence, either conscious or unconscious, received from and exerted by agents participating in the experiment. Random sampling, on the other hand, is intended to justify, somehow, assumptions concerning the functional form of a distribution in the statistical model of the experiment. The distinction between random design and random sampling will be kept here, even though, as briefly mentioned in section 6, a deeper probabilistic analysis of randomization shows that, from a theoretical point of view, the two concepts can greatly overlap.

Our immediate interest in randomization (and control) is on whether it can assist the design of experiments by inducing independence relations. This strategy is pinpointed in the following quote from Pearl's epilogue: "The Art and Science of Cause and Effect":

> Fisher's 'randomized experiment'... consists of two parts, 'randomization' and intervention.' (Pearl, 2000, p. 340)

> Intervention means that we change the natural behavior of the individual: we separate subjects into two groups, called treatment and control, and we convince the subjects to obey the experimental policy. We assign treatment to some patients who, under normal circumstances, will not seek treatment, and give placebo to patients who otherwise would receive treatment. That, in our new vocabulary, means 'surgery' -- we are severing one functional link and replacing it with another. Fisher's great insight was that connecting the new link to a random coin flip 'guarantees' that the link we wish to break is actually broken. The reason is that a random coin is assumed to be unaffected by anything we can measure on macroscopic level... (Pearl, 2000, p. 348)

## 5 C. S. Peirce and Randomization

We believe that many fine points about the role of randomization in the DSEs can be better understood by following its development from an historical perspective. This is the topic of this section.

In the period of 1850 to 1880 the quantitative analysis of human sensation in response to physical (tactile, acoustic or visual) stimuli, was the main goal of psychophysics. A typical hypothesis in this research program was Fechner's law, see Hernstein and Boring (1966, p. 72), which stated that,

> The magnitude of sensation ($\gamma$) is not proportional to the absolute value of the stimulus ($\beta$), but rather to the logarithm of the magnitude of the stimulus when this is expresses in terms of its threshold value ($b$), i.e. that magnitude considered as unit at which the sensation begins and disappears."

In modern mathematical notation, $\gamma = k \log(\beta/b)\ I(\beta > b)$ .

In his psychophysical experiments Fechner tested his own ability to distinguish the strongest in a pair of stimuli. For example, he would prepare two objects of masses $\mu$ and $\mu+\delta$, and later on he would lift them, and "answer" which one appeared to him to be the heaviest. A quantitative analysis would latter relate the proportion of right and wrong answers with the values of $\mu$ and $\delta$, see Stigler (1986, ch. 7, Psychophysics as a Counterpoint, pp. 239-261). Fechner was well aware of the potential difficulties resulting from the fact that the experiments where not performed blindly, that is, since he prepared the experiment himself, he could know in advance the right answer. Nevertheless, he claimed to be able to control himself, be objective, and overcome this difficulty.

According to Dehue (1997), in the decade of 1870, G. E. Müller and several researchers at Tübingen and Göttingen Universities, began to improve the design of psychophysical experiments. The first major improvement was blinding: the stimuli were prepared or administered by an "Experimenter" or "Operator" and applied to a distinct person, the "Observer", "Patient" or "Subject", who was kept unaware of the actual intensity of the stimuli.

The second major improvement was the precaution of presenting the stimuli in irregular order (buntem Wechsel). This irregularity was introduced to prevent the patient from becoming habituated to patterns in the sequence of stimuli presented to him or, in other words, to keep him to form building expectations and guessing the right answers. Nevertheless, there was, at that time, neither a general theory defining *irregularity*, nor a systematic method for providing an irregular order.

In 1885, Charles Sanders Peirce and his student Joseph Jastrow presented randomization as a practical solution, in this context, to the question of irregularity, that is, systematic randomization should prevent any effective guessing by the patient, see Hacking (1988, pp. 431-434). Peirce was in fact insisting on *exchangeability*, a key notion in the analysis of randomization in modern statistics and, most specially, in Bayesian statistics, that will be discussed in the next section.

Peirce also struggled with the dilemma of allowing or not, in the course of the experiment, sequences that do not appear to be random. His conclusions, see Peirce and Jastrow (1884, p. 122), are, once more, precursors to De Finetti's concept of exchangeability:

> The pack (of playing-cards) was well shuffled, and, the operator and subject having taken their places, the operator was governed by the color of the successive cards ...
>
> A slight disadvantage in this mode of proceeding arises from the long runs of one particular kind of change, which would occasionally be produced by chance and would tend to confuse the mind of the subject. But it seems clear that this disadvantage was less than that which would have been occasioned by his knowing that there would be no such long runs if any means had been taken to prevent them. (Peirce & Jastrow, 1884, p. 122)

Regardless of its importance, Peirce's solution of randomization was not accepted by his contemporaries, fell into oblivion, and was almost forgotten, until it reappeared much later in the work of R. A. Fisher. We believe that there are several entangled reasons to explain such a twisted historical process. The psychophysics community raised objections against some of the hypotheses, and also against some methodological aspects presented in Peirce's paper. Besides, there is also a confounding factor generated by a second role played by randomization in Peirce's paper, namely, randomization to measure faint effects. We shall briefly discuss these aspects in the next paragraphs.

Fechner assumed the existence of a threshold (Schwelle), $b$, below which small differences could no longer be discerned. Peirce wanted to refute the existence of this threshold assuming, instead, a continuously decreasing sensitivity to smaller and smaller differences. We should remark that for Peirce this should not have been a fortuitous hypothesis, since it can be related to his general philosophical ideas, most specially with the concept of synechism, see Hartshorne et al. (1992), Eisele (1976), and Stern (2007b).

Peirce postulated that the patients' sensitivity could be adequately measured by the probability of correct answers, even when the difference was too faint to be consciously discerned by the same patients. Hence, in experiments similar to Fechner's, Peirce asked the patient always to guess the correct answer. Peirce also asked the patient to give the answer a confidence score from 0 to 3. Peirce analyzed his experimental data and derived empirical formulae relating the (rounded) "subjective" confidence scores, $m$, and the "objective" probability of correct answers, $p$:

> The average marks seem to conform to the formula $m=c\log(p/(1\text{-}p))$, where $m$ denotes the degree of confidence on the scale, $p$ denotes the probability of the answer being right, and $c$ is a constant which may be called the index of confidence. (Peirce & Jastrow, 1884, p. 122)

At the time of Peirce's experiments, the psychophysical community gave great importance to the analysis of the patient's subjective 'introspections'. According to this view, Peirce's experiments were criticized by asking the patient to guess the correct answer even when he expressed low confidence. Of course, if one understands Peirce's research program, it is clear that that the experimental design he used is perfectly coherent. Unfortunately, this was not the judgment of his contemporaries.

The same techniques and experimental designs used by Peirce were subsequently used by several researchers in attempts to measure faint effects, including effects produced by 'below the consciousness threshold', sub-conscious, or sub-liminal stimuli. Some of these studies were really misconceived, and that may have been yet another contributing factor for the reactions against the use of randomization. Whatever the explanation might be, Peirce's paper fell into oblivion, and the progress of DSEs was delayed by half a century.

## 6 Bayesian Analysis of Randomization

The work of Ronald Aylmer Fisher can undoubtedly be held responsible for disseminating the modern approach to DSEs, including randomization, to almost any area of empirical research, see for example Fisher (1926, 1966). The idea of randomization, however, was later contested by some members of the Bayesian school. Commenting on the use of randomization after Fisher, Hacking states:

> Undoubtedly Fisher won the day, at least for the following generation, but then a new, although not completely unrelated, challenge to randomized design arose. This came from the revival of the

> 'Bayesian' school, typically associated with L. J. Savage's theory of what he called personal probability. Here the object is to form an initial assessment of one's personal beliefs about a subject and to modify them in the light of experience and a theoretical analysis formally modeled by the calculus of probability and a theory of personal utility. It is widely held to be an almost immediate consequence of this approach that randomization is of no value at all (except perhaps to eliminate some kind of fraud. (Hacking, 1988, pp. 429-430)

This erroneous notion of incompatibility between the use of randomization and Bayesian statistics in now completely outdated. One of the most prestigious textbooks in contemporary Bayesian statistics states:

> A naive student of Bayesian inference might claim that because all inference is conditional on the observed data, it makes no difference how those data were collected. This misplaced appeal to the likelihood principle would assert that given (1) a fixed model (including the prior distribution) for the underlying data and (2) fixed observed values of the data, Bayesian inference is determined regardless of the design for the collection of the data. Under this view there would be no formal role for randomization in either sample surveys or experiments. The essential flaw in the argument is that a complete definition of 'the observed data' should include information on how the observed values arose, and in many situations such information has a direct bearing on how these values should be interpreted. Formally then, the data analyst needs to incorporate the information describing the data collection process in the probability model used for analysis. (Gelman et al., 2003, ch.7, p. 198)

Indeed, the classical argument using the likelihood principle against randomization in the DSEs, assumes a fixed, given statistical model and, as concisely stated by Kempthorne:

> The assertion that one does not need randomization in the context of the assumed (linear) model (above) is an empty one because an intrinsic role of randomization is to 'insure' against model inadequacies. (Kempthorne, 1977, p. 16)

Gelman et al. (2003) proceeds offering a much deeper analysis of the role of randomization from a Bayesian perspective, see also Rubin (1978). The key concept of "ignorable design" specifies decoupling conditions between the sampling (or censoring) process, described by an indicator variable, $I$, and the distribution of the observed variables, $y_{obs}$. If the experiment has an ignorable design, we can build a statistical model that explicitly considers $y_{obs}$ alone. Finally, it is ironic that perhaps one of the best arguments for incorporating randomization in Bayesian experimental design is a consequence of the de Finetti theorem for exchangeability. As mentioned in section 4, this argument also blurs the distinction between the concepts of randomized design and randomized sampling. We quote, once again, from Gelman et al.:

> How does randomization fit into this picture? First, consider the situation with no fully observed covariates $x$, in which case the 'only' way to have an invariant to permutation design -- is to randomize.
>
> In this sense, there is a benefit to using different patterns of treatment assignment for different experiments; if nothing else about the experiments is specified, they are exchangeable, and the global treatment assignment is necessarily randomized over the set of experiments. (2003, pp. 223-225)

## 7 Randomization, Epistemic Considerations

Several researchers currently concerned with epistemological questions in Bayesian statistics are engaged in a reductionist program dedicated to translate every statistical test or inference problem into a decision theoretic procedure. One of the main proponents and early contributors to this program, but one who also had a much broader perspective, clearly articulating his epistemological insights and motivations, was Bruno de Finetti.

In statistical models our knowledge of the world is encoded in probability distributions. Hence, it is vital to clarify the epistemological or ontological status of probability. Let us examine de Finetti's position, based on his own words:

> Any assertion concerning probabilities of events is merely the expression of somebody's opinion and not itself an event. There is no meaning, therefore, in asking whether such an assertion is true or false, or more or less probable." (de Finetti, 1972, p. 189)

> Each individual making a 'coherent' evaluation of probability (in the sense I shall define later) and desiring it to be 'objectively exact', does not hurt anyone: everyone will agree that this is his subjective evaluation and his 'objectivist' statement will be a harmless boast in the eyes of the subjectivist, while it will be judged as true or false by the objectivist who agree with it or who, on the other hand, had a different one. This is a general fact, which is obvious but insignificant: 'Each in his own way.' (de Finetti, 1980, p. 212)

Solipsism, from the Latin solus (alone) + ipse (self), can be defined as the epistemological thesis that the individual's subjective states of mind are the only proper or possible basis of knowledge. Metaphysical solipsism goes even further, stating that nothing really 'exists' outside of one's own mind. From the two above quotations, it is clear that de Finetti stands, if not from a metaphysical, at least from a epistemological perspective, as a true solipsist. This goes farther than many theorists of the Bayesian subjectivist school would venture, but de Finetti charges ahead, with a program that is not only anti-realist, but also anti-idealist. De Finetti launches a full-fledged attack against the vain and futile desire for any objective knowledge:

> Much more serious is the reluctance to abandon the inveterate tendency of the savages to objectivize and mythologize everything (1); a tendency that, unfortunately, has been, and is, favored by many more philosophers than have struggled to free us from it (2).
> (1) The main responsibility for the objectivizationistic fetters inflicted on thought by everyday language rests with the verb 'to be' or 'to exist', and this is why we drew attention to it in the exemplifying sentences. From it derives the swarm of pseudoproblems from 'to be or not to be', to 'cogito ergo sum', from the existence of 'cosmic ether' to that of 'philosophical dogmas'.
> (2) This is what distinguishes acute minds, who enlivened thought and stimulated its progress, from narrow-minded spirits who mortified and tried to mummify it ... 'great thinkers' (like Socrates and Hume) and 'school philosophers' (like Plato and Kant). (Finetti, 1974, vol. 1, Sec. 1.11, pp. 21-22)

De Finetti was also aware of the dangers of *objective contamination*, that is, any "objective" (probabilistic) statement can potentially "infect" and spread its objectivity to other statements:

> There is no way, however, in which the individual can avoid the burden of his own evaluations. The key can not be found that will unlock the enchanted garden wherein, among the fairy-rings and the shrubs of magic wands, beneath the trees laden with monads and noumena, blossom forth the flowers of 'Probabilitas realis'. With the fabulous blooms safely in our button-holes we would be spared the necessity of forming opinions, and the heavy loads we bear upon our necks would be rendered superfluous once and for all. (Finetti, 1974, vol. 2, Sec. 7.5.7, pp. 41-42)

As we have seen in the last sections, a randomization device is built so to provide legitimate 'objective' probabilistic statements about some events, and randomization procedures in DSEs are conceived exactly in order to spread this objectivity around.

I. J. Good was an other leading figure of the early days of the Bayesian revival movement. Contrary to de Finetti, Good has always been aware of the dangers of an extreme subjectivist position:

> Some of you might have expected me, as a confirmed Bayesian, to restrict the meaning of the word 'probability' to subjective (personal) probability. That I have not done so is because I tend to believe that physical probability exists and is in any case a useful concept. I think physical probability can be measured only with the help of subjective probability, whereas de Finetti believes that it can be 'defined' in terms of subjective probability. De Finetti showed that if a person has a consistent set of subjective or logical probabilities, then he will behave 'as if' there were physical probabilities, where the physical probability has an initial subjective probability distribution. It seems to me that, if we are going to act if the physical probability exists, then we don't lose anything practical if we assume it really 'does' exist. In fact I am not sure that existence means more than there are no conceivable circumstances in which the assumption of existence would be misleading. But this is perhaps too glib a definition. The philosophical impact of de Finetti's theorem is that it supports the view that solipsism cannot be logically disproved. Perhaps it is the mathematical theorem with most potential philosophical impact. (Good, 1983, p. 93)

In our terminology we would have used the expression 'objective probability' instead of Good's expression, 'physical probability'. In 1962 Good edited a collection of speculative essays, including some on the foundations of statistics. The following short essay by Christopher S.O'D. Scott offers an almost direct answer to de Finetti:

> Scientific Inference: You are given a large number of identical inscrutable boxes. You are to select one, the 'target box', by any means you wish which does not involve opening any boxes, and you then have to say something about is in it. You may do this by any means you wish which does not involve opening the target box.
>
> This apparent miracle can easily be performed. You only have to select the target box at random, and then open a random sample of other boxes. The contents of the sample boxes enable you to make an estimate of the contents of the target box which will be better than a chance guess. To take an extreme case, if none of the sample boxes contains a rabbit and your sample is large, you can state with considerable confidence: 'The target box does not contain a rabbit.' In saying this, you make no assumption whatever about the principles which may have been used in filling the boxes.
>
> This process epitomizes scientific induction at its simplest, which is the basis of all scientific inference. It depends only on the existence of a method of randomization – that is, on the assumption that events can be found which are unrelated (or almost) to given events.

> It is usually thought that scientific inference depends upon nature being orderly. The above shows that a seemingly weaker condition will suffice: Scientific inference depends upon our knowing ways in which nature is disorderly. (Good, 1962, sec.114, pp. 364-365)

In the preceding articles we discussed general conditions validating objective knowledge, from a constructivist epistemological perspective. In this article we discuss the use of randomization devices, that can generate observable events with distribution that are independent of the distribution of any event relevant to a given statistical study. For example, the statistical study could be concerned with the reaction of human patients affected by a given disease to alternative medical treatments, whereas a good randomization device could be a generic "coin flipping machine," like a regular dice or a mechanical roulette borrowed from a casino. The randomization device could also be a sophisticated apparatus detecting flips (state transitions) in some quantum system, with transitions probabilities known with a relative precision of one over a trillion.

So far in this article we have seen how well can decoupling strategies used in the DSEs, including randomization procedures, help us to perform robust statistical inference and, in doing so, escape, from a pragmatic perspective, the solipsist burdens of an extreme subjectivist position. The same techniques can induce no association relations, generating sparse or structured statistical models. No association hypotheses can then be tested, confirming (or not) such sparse or structured patterns in the statistical model.

## 8 Final Remarks and Further Research

As analyzed in this article, the randomization method, introduced by C. S. Peirce and J. Jastrow (1884), is the fundamental decoupling technique used in the design of statistical experiments (DSEs). Nevertheless, only after the work of R. A. Fisher (1966), were randomized designs used regularly in practice. Today, randomization is one of the basic backbones of statistical theory and methods. Meanwhile, the pioneering work of Peirce had been virtually forgotten by the statistics community, until rediscovered by the historical research of Stigler (1978) and Hacking (1988). Nevertheless, even today, the work of Peirce is presented as an isolated and ad hoc contribution. As briefly indicated in section 5, it is plausible that Peirce and Jastrow's experimental and methodological work could have had motivations related to more general ideas of Peircean philosophy. In particular, we believe that the faint effects psychophysical hypothesis can be liked to the concept of synechism, while the randomized design solution can be embedded in the epistemological framework of Peirce's objective idealism. We believe that these topics deserve the attention of further research.

In this article we have examined some aspects of DSEs, such as blocking, control and randomization, from an epistemological perspective. However, in many applications, most noticeably in medical studies, several other aspects have to be taken into account, including the well being of the patients taking part in the study. In our view, such complex situations require a thorough, open and honest discussion of all the moral and ethical aspects involved. Typically they also demand sound protocols and complex statistical models, suited to the fine quantitative analyses needed to balance multiple objectives and competing goals. For the placebo, nocebo, kluge Hans [Clever Hans], and similar effects, and the importance of blind and randomized clinical trials, see Kotz et al. (2005), under the entries Clinical Trials I, by N. E. Breslow, vol. 2, pp. 981-989, and Clinical Trials II, by R. Simon, vol. 2, pp. 989-998. For additional references on statistical randomization procedures, see Folks (1984), Kadane and Seidenfeld (1990), Kaptchuk and Kerr (2004), Karlowski et al. (1975), Kempthorne (1977, 1980) and Noseworthy et al. (1994).

## References


Benade, A. H. (1992). *Horns, strings, and harmony.* Mineola, NY: Dover.

Bertsekas, D. P., Tsitsiklis, J. N. (1989). *Parallel and distributed computation, numerical methods*. New York: Prentice Hall.

Blair, J. R. S., Peyton, B. (1993). An introduction to chordal graphs and clique trees. In George, A., Gilbert, J. R., Liu, J. W. H. (Eds.), *Graph theory and sparse matrix computation* (pp. 1-30). New York: Springer.

Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, *67* (338), 364-366.

Box, G. E. P., Hunter, W. G., Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design, data analysis and model building*. New York: Wiley.

Bunch, J. R., Rose, D. J. (1976). *Sparse matrix computations*. New York: Academic Press.

Butkov, E. (1968). *Mathematical physics*. Reading, MA: Addison-Wesley.

Censor, Y., Zenios, S. A. (1998). *Parallel optimization: Theory, algorithms, and applications*. New York: Oxford University Press.

Cichocki, A., Amari, S. I., Zdunek, R., Kompass, R., Hori, G., He, Z. (2006). Extended SMART algorithms for non-negative matrix factorization. *Lecture Notes in Artificial Intelligence*, *4029*, 548-562.

Cichocki, A., Zdunek, R., Amari, S. I. (2006). Csiszar's divergences for non-negative matrix factorization: Family of new algorithms. *Lecture Notes in Computer Science*, *3889*, 32-39.

Cobb, G. W. (1998). *Introduction to design and analysis of experiments*. New York: Springer.

Cozman, F. G. (2000). Generalizing variable elimination in Bayesian networks. In the *Proceedings of PRAI 2000, Workshop on Probabilistic Reasoning in Artificial Intelligence* (pp. 27-32). IBERAMIA-SBIA 2000, 7th Ibero-American Conference on Artificial Intelligence and 15th Brazilian Conference on Artificial Intelligence. Atibaia, Brazil.

Dehue, T. (1997). Deception, efficiency, and random groups: Psychology and the gradual origination of the random group design. *Isis*, *88* (4), 653-673.

Dhillon, I. S., Sra, S. (2005). Generalized nonnegative matrix approximations with Bregman divergences. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (pp. 283-290). Cambridge, MA: Thi MIT Press.

Eisele, C. (Ed). (1976). *The new elements of mathematics of Charles S. Peirce*. The Hague: Mouton.

Finetti, B. de (1974). *Theory of probability,* vol. 1 and vol. 2. London: Wiley.

Finetti, B. de (1972). *Probability, induction and statistics: The art of guessing*. London: John Wiley.

Finetti, B. de (1980). Probability: Beware of falsifications. In H. Kyburg & H. E. Smokler (Eds.), *Studies in subjective probability* (pp. 193-224). New York: Krieger.

Fisher, R. A. (1966). *The design of experiments*. (8th ed.) London: Oliver and Boyd. (1st ed., 1935)

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture*, *33*, 503-513.

Folks, J. L. (1984). Use of randomization in experimental research. in K. Hinkelmann, *Experimental design, statistical models and genetic statistics*. *Essays in honor of Oscar Kempthorne* (pp. 17–32). Basel, Switzerland: Marcel Dekker.

Franklin, J. N. (1968). *Matrix theory.* Englewood Cliffs, NJ: Prentice-Hall.

French, A. P. (1974). *Vibrations and waves*. M.I.T. Introductory Physics Series. Cambridge, MA: The MIT Press.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2003). *Bayesian data analysis*. New York: Chapman and Hall/CRC.

Geoffrion, A. M. (Ed.) (1972). *Perspectives on optimization: A collection of expository articles*. New York: Addison-Wesley.

George, A., Liu, J. W. H. (1978). A quotient graph model for symmetric factorization. In I. S. Duff & G. W. Stewart (Eds.), *Sparse matrix proceedings* (pp. 154-175). Philadelphia: SIAM.

George, A., Liu, J. W. H., Ng, E. (1990). Solution of sparse positive definite systems on a hypercube. In H.A. van der Vorst, P. van Dooren, (Eds.), *Parallel algorithms for numerical linear algebra* (pp.129-156). Amsterdam: North-Holland.

George, A., Gilbert, J. R., Liu, J. W. H. (Eds.) (1993). *Graph theory and sparse matrix computation*. New York: Springer.

George, A., Liu, J. W. H. (1981). *Computer solution of large sparse positive-definite systems*. New York: Prentice-Hall.

Golub, G. H., & Loan, C. F. van (1989). *Matrix computations*. Baltimore: Johns Hopkins University Press.

Golumbic, M. C. (1980). *Algorithmic graph theory and perfect graphs*. New York: Academic Press.

Good, I. J. (Ed.) (1962). *The scientist speculates. An anthology of partly-baked ideas*. New York: Basic Books.

Good, I. J. (1983). *Good thinking*. Minneapolis: Univ. of Minnesota.Press.

Good, I. J., Mittal, Y. (1987). The amalgamation and geometry of two-by-two contingency Tables. *Annals of Statistics*, *15*, pp. 694-711.

Gotzsche, P. C. (2006). Assessment of bias. In S. Kotz (Ed.), *The Encyclopedia of Statistics, 1*, 237-240.

Greenland, S., Pearl, J., Robins, J. M. (1999). Confounding and collapsibility in causal inference. *Statistical Science 14* (1), 29-46.

Greenland, S., Morgenstern, H. (2001). Confounding in health research. *Annual Review of Public Health*, *22*, 189-212.

Hacking, I. (1988). Telepathy: Origins of randomization in experimental design. *Isis*, *79* (3), 427-451.

Hartshorne, C., Weiss, P., Burks, A. (Eds.) (1992). *Collected papers of Charles Sanders Peirce*. Charlottesville: InteLex Corp.

Haupt, E. J. (1998). G. E. Müller as a source of American psychology. In R. W. Rieber, K. Salzinger, (Eds.),. *Psychology: Theoretical-historical perspectives* (pp. 16-75). Washington D.C.: American Psychological Association.

Hernstein, R. J., Boring, E. G. (1966). *A source book in psychology.* Cambridge, MA: Harvard University Press]

Heydtmann, M. (2002). The nature of truth: Simpson's paradox and the limits of statistical data. *QJM: An International Journal of Medicine*, *95* (4), 247-249.

Hinkelmann, K. (Ed.) (1984). *Experimental design, statistical models and genetic statistics. Essays in honor of Oscar Kempthorne*. Basel, Switzerland: Marcel Dekker.

Hocking, R. R. (1985). *The analysis of linear models*. Monterrey, CA: Brooks Cole.

Hoyer, P. O. (2004). Non-negative matrix factorizations with sparseness constrains. *J. of Machine Learning Research*, *5*, 1457-1469.

Jastrow, J. (1988). A critique of psycho-physic methods. *American Journal of Psychology*, *1*, 271-309.

Kadane, Y, Seidenfeld, T. (1990). Randomization in a Bayesian perspective. *J. of Statistical Planning and Inference*, *25*, 329-345.

Kaptchuk, T. J., Kerr, C. E. (2004). Commentary: Unbiased divination, unbiased evidence, and the Patulin clinical trial. *International Journal of Epidemiology*, *33*, 247-251.

Karlowski, T. R., Chalmers, T. C., Frankel, T. C., Kapikian, L. D. Lewis, T. L., Lynch, J. M. (1975). Ascorbic acid for the common cold: a prophylactic and therapeutic trial. *JAMA*, *231*, 1038-1042.

Kemptorne, O. (1980). Foundations of statistical thinking and reasoning. *Australian CSIRO-DMS Newsletter, 68*, 1–5; *69*, 3–7.

Kempthorne, O. (1977). Why randomize? *J. of Statistical Planning and Inference, 1*, 1-25

Kloesel, C. J. W. (Ed.) (1993). *Writings of Charles S. Peirce. A chronological edition*, vol.5, 1884-1886.

Kotz, S., Balakrishnan, N., Read, C.B., Vidakovic, B. (Eds.). (2005). *Encyclopedia of statistical sciences* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Lauretto, M., Pereira, C. A. B., Stern, J. M., Zacks, S. (2003). Full Bayesian significance test applied to multivariate normal structure models. *Brazilian Journal of Probability and Statistics, 17*, 147-168.

Lauritzen, S. L. (2006). *Fundamentals of graphical models*. Saint Flour Summer-school Lecture Notes. Retrieved on September 13, 2006 from http://www.stats.ox.ac.uk/steffen/stflour/

Lasdon, L. S. (1970). *Optimization theory for large systems*. New York: MacMillan.

Lindley, D. V. (1991). *Making decisions*. New York: John Wiley.

Lindley, D. V., Novick, M. R. (1981). The role of exchangeability in inference. *The Annals of Statistics*, *9* (1), 45-58.

Marion, J. B. (1975). *Classical dynamics of particles and systems*. New York: Academic Press.

Noseworthy, J. H., Ebers, G. C., Vandervoort, M. K., Farquhar, R. E., Yetisir, E., Roberts, R. (1994). The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology, 44*, 16-20.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J. (2004). *Simpson's paradox: An anatomy.* Research Report, Cognitive Systems Lab., Computer Science Dept., Univ.of California at Los Angeles.

Peirce, C. S., Jastrow, J. (1884). On small differences of sensation. *Memoirs of the National Academy of Sciences*, 3 (1884), 75-83. Also in Kloesel (1993), Vol. 5 (1884-1886), pp. 122-135.

Pissanetzky, S. (1984). *Sparse matrix technology.* New York: Academic Press.

Reintjes, R., Boer, A. de, Pelt, W. van, Groot, J. M. de (2000). Simpson's paradox: An example from hospital epidemiology. *Epidemiology*, *11*, 1, 81-83.

Rijsbergen, C. J. van (2004). *The geometry of information retrieval*. Cambridge: Cambridge University Press.

Rose, D. J. (1972) *Sparse matrices and their applications*. New York: Springer.

Rose, D. J., Willoughby, R. A. (1972). *Sparse matrices*. New York: Plenum Press.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics, 6*, 34-58.
Sadun, L. (2001). *Applied linear algebra: The decoupling principle*. New York: Prentice Hall.
Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, Ser. B, *13*, 238-241.
Schlick, M. (1920). Naturphilosophische Betrachtungen über das Kausalprintzip. *Die Naturwissenschaften*, 8, 461-474. Translated as, Philosophical Reflections on the Causal Principle, ch. 12, pp. 295-321, v.1 in M.Schlick (1979).
Schlick, M. (1979). *Philosophical papers*. Dordrecht: Reidel.
Stern, J. M. (1992). Simulated annealing with a temperature dependent penalty function. *ORSA Journal on Computing*, *4* (3), 311-319.
Stern, J. M. (1994). *Esparsidade, Estrutura, Estabilidade e Escalonamento em Álgebra Linear Computacional*. Recife, Brazil: IX Escola de Computação.
Stern, J. M. (2006a). Decoupling, sparsity, randomization, and objective Bayesian inference. Tech. Rep. MAC-IME-USP-2006-07, University of São Paulo.
Stern, J. M. (2006b). Language, metaphor and metaphysics: The subjective side of science. Tech. Rep. MAC-IME-USP-2006-09, University of São Paulo.
Stern, J. M. (2007a). Cognitive constructivism, eigen–solutions, and sharp statistical hypotheses. *Cybernetics and Human Knowing*, *14* (1), 9-36.
Stern, J. M. (2007b). Language and the self-reference paradox. *Cybernetics and Human Knowing*, *14* (4), 71-92.
Stern, J. M., Vavasis, S. A. (1993). Nested dissection for sparse nullspace bases. *SIAM Journal on Matrix Analysis and Applications*, *14* (3), 766-775.
Stern, J. M., Vavasis, S.A. (1994). Active set algorithms for problems in block angular form. *Computational and Applied Mathematics*, *12* (3), 199-226.
Stigler, S. M. (1978). Mathematical statistics in the early states. *The Annals of Statistics, 6*, 239-265.
Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard Univ. Press.
Vorst, H. A. van der, Dooren, P. van (Eds.) (1990). *Parallel algorithms for numerical linear algebra*. Amsterdam: North-Holland.
Wismer, D. A. (Ed.) (1971). *Optimization methods for large-scale systems with applications*. New York: McGaw-Hill.

Taylor, C. (2003). *Magma III* (rotated). 40” x 72”, oil on canvas.