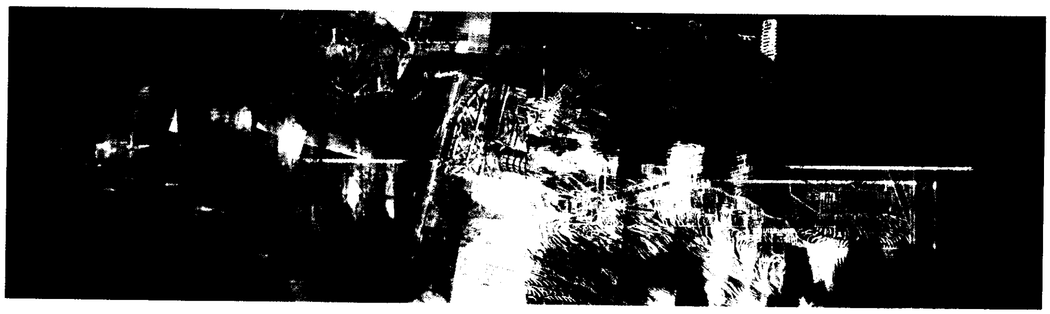
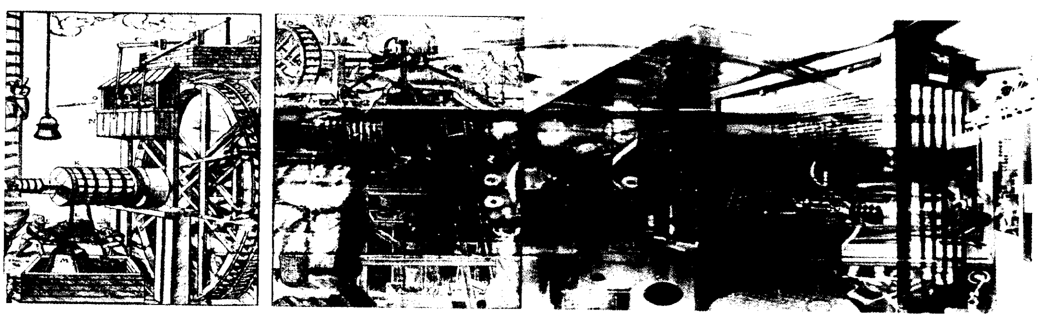
offprint

CYBERNETICS & HUMAN KNOWING

a journal of second-order cybernetics
autopoiesis and cyber-semiotics
Volume 22, No. 1, 2015





What Underlies Information?

Contents

The artist for this issue is Bill Seaman
Cover art: Seaman, Bill (2014). Engine of Many Senses.
Generative video/audio/text installation; stills B 07, 14, 02

CYBERNETICS & HUMAN KNOWING

A Journal of Second-Order Cybernetics, Autopoiesis & Cyber-Semiotics

ISSN: 0907-0877

Cybernetics and Human Knowing is a quarterly international multi- and trans-disciplinary journal focusing on second-order cybernetics and cybersemiotic approaches.

The journal is devoted to the new understandings of the self-organizing processes of information in human knowing that have arisen through the cybernetics of cybernetics, or second order cybernetics its relation and relevance to other interdisciplinary approaches such as C.S. Peirce's semiotics. This new development within the area of knowledge-directed processes is a non-disciplinary approach. Through the concept of self-reference it explores: cognition, communication and languaging in all of its manifestations; our understanding of organization and information in human, artificial and natural systems; and our understanding of understanding within the natural and social sciences, humanities, information and library science, and in social practices like design, education, organization, teaching, therapy, art, management and politics. Because of the interdisciplinary character articles are written in such a way that people from other domains can understand them. Articles from practitioners will be accepted in a special section. All articles are peer-reviewed.

Subscription Information

Price: Individual £66. Institutional: £145 (online); £178 (online & print). 50% discount on full set of back volumes. Payment by cheque in £UK (pay Imprint Academic) to PO Box 200, Exeter EX5 5HY, UK; Visa/Mastercard/Amex.
email: sandra@imprint.co.uks

Editor in Chief: Søren Brier, Professor in semiotics at the Department of International Culture and Communication Studies attached to the Centre for Language, Cognition, and Mentality, Copenhagen Business School, Dalgas Have 15, DK-2000 Frederiksberg, Denmark, Tel: +45 38153246. sb.ikk@cbs.dk

Editor: Jeanette Bopry, Instructional Sciences, National Institute of Education, 1 Nanyang Walk, Singapore 637616. jeanette.bopry@gmail.com

Associate editor: Dr. Paul Cobley, Reader in Communications, London Metropolitan University, 31 Jewry Street, London EC3N 2EY. p.cobley@londonmet.ac.uk

Managing editor: Phillip Guddemi, The Union Institute and University, Sacramento CA, USA. pguddemi@well.com

Joint art and website editor: Claudia Jacques
cj@claudiajacques.org

C&HK is indexed/abstracted in *Cabell's Journal* and *PsycInfo*
Journal homepage: www.chkjournal.org
Full text: www.ingenta.com/journals/browse/imp

Cybernetics and Human Knowing. Vol. 22 (2015), no. 1, pp. 9-33

Peirce and Spencer-Brown on Probability, Chance, and Lawfulness

John Levi Martin[1]

Before the pivotal work *The Laws of Form,* which made him influential among systems theorists, George Spencer-Brown had achieved wide publicity for work on statistics that seemed to explain away accumulated findings for extra-sensory perception. Interestingly, just as in his later work, there was a remarkable convergence here with the earlier writings of C. S. Peirce. Both emphasized the difference between the randomness of generating processes and the empirical distributions used for the production of generalizations. Both understood this as challenging theories of scientific inference. Yet Spencer-Brown's conclusion—that science and magic were both eaten away by the tides of time through the accumulation of patterns through randomness—was not a necessary one for Peirce, for whom these patterns might have ontological significance, as they wore grooves of habit into the universe. Grappling with the puzzles at the heart of these questions may change how we incorporate the notion of information into our theories.
Keywords: Probability, Chance, Peirce, Spencer-Brown, Law, Magic, ESP

Here I wish to give some attention to an issue that might seem extremely minor in the history of social thought, namely the interpretation of probabilities. And I will do so by focusing on the surprising re-discovery of points first made by Charles Sanders Peirce, the founder of American pragmatism, by George Spencer-Brown, the logician whose work inspires much of German systems theory, and in particular, the work of Niklas Luhmann and Dirk Baecker. I will give relatively close attention to previously overlooked parallels in their critiques of probability. I then consider the implications of the Peircean/Spencer-Brownian critique for our notion of information, and our ideas of the approach to entropy that has been used in information theory to quantify this.

To some extent, it may be significant in itself that the convergence between the two, noted in their mathematical work, is mirrored in an independent convergence in their approach to probability. Whether this involved actual influence is a matter left for detailed historical research. But I believe that there is a more practical issue at hand: Most social thought that uses the concept of information assumes a sufficiently large sample in all cases such that there is little interest in probing the difference between the probabilities of a generating process, and the distribution of realizations of this process. In a nutshell, *inference* is taken to be completely unproblematic. Both Peirce and Spencer-Brown, in contrast, focused their attention on the necessary slippage between these two, and, in rethinking the issue of inference, challenged our understandings of what science is, and what it can be.

1. Department of Sociology, University of Chicago Chicago, IL. Email: jlmartin@uchicago.edu

I begin by discussing the general problem of inference as it posed itself to most mathematical scientists, and then examine how both Peirce and Spencer-Brown linked this to the issue of probability. After exploring the relationship between chance and time, we return to the issue of attempting to infer lawfulness from data, and then use this to rethink the issues of *expectation* that are central for information theory. Throughout, I will return to Spencer-Brown's emphasis on the puzzle of deciding what is, and what is not, a chance machine—which turns out to be critical for any attempt to found a social theory on information.

Peirce, Spencer-Brown, and Other Spirits

Improbable Overlaps

C. S. Peirce and Spencer-Brown, so far as one can tell, shared a great deal in common. Both were brilliant, rambunctious and sometimes a bit too fond of irritating others; both had difficulties in pursuing a conventional academic career, due to erratic work habits, their pursuit topics whose relevance most peers did not accept, and perhaps an overactive physical libido.[2] Further, both considered themselves to have (in Spencer-Brown's [2004, p. 111] words), an "enormous Sherlock-Holmes capacity" (on Peirce and Holmes see Sebeok and Umiker-Sebeok [1983] as well as other pieces in this collection edited with Umberto Eco; for Eco's own construction of this method in a fictional character see Eco [1980/1983, 304f]).

Charles Sanders Peirce was the son of the leading mathematician of his era, and his father would consistently step in to save him whenever he overextended his financial or social resources. He identified mainly as a logician, though it was his philosophy of science that first attracted attention; he is now mainly prized for his invention of semiotics, but many of his innovations in logic—perhaps notably, his invention of lattice theory—were independently rediscovered by others (in the case of lattice theory, by Garrett Birkhoff[3]).

Spencer-Brown pursued a conventional training in mathematics, along with a wide variety of other pastimes; like Peirce, he supported himself afterwards in applied science and seems to have taken seriously his work in elevator systems (while Peirce's dealt with pendulums). But he also made an entrance to the circle of young logicians surrounding Bertrand Russell. Russell was, unlike most of his peers, extremely happy

2. In his autobiography begun when he was around 80, Spencer-Brown (2004, p. 47) wonderfully used his early childhood recollections as a substitute for Facebook, still hoping to re-connect with some of the darlings who perhaps first inducted him into the hall of fame of love; to be fair, he also reached out to all the potential tathagata who might inhabit human form (1995, p. 142 n22).
3. Birkhoff (1940/1979) was aware of and influenced by some of Peirce's early published work on relational logic.

to foster a younger generation of mathematicians who were bent on destroying his own life's work—Gödel, Wittgenstein, and Spencer-Brown himself.[4]

It was this critique of the Russell-Whitehead system that led to Spencer-Brown's *Laws of Form*, which became a cult favorite, highly influential on those working in the creative interstitial areas around systems theory (and the relation of these to psychology, as with R. D. Laing, Spencer-Brown's adopted teacher and therapist at this point, or the somewhat more academically respectable Gregory Bateson). In recent years, most importantly, in a special issue of the journal *Cybernetics and Human Knowing*, there has been some attention to the remarkable convergence between Spencer-Brown's work here and earlier efforts by Peirce. (For other discussions of important parallels between Peirce and Spencer-Brown see Brier [2014, p. 7f].) Most obviously, both developed a system that turned on a single sign combining negation (as indicated by an overbar) and the Sheffer stroke | which indicates neither-nor (see Kauffman [2001, p. 80f] on the relation between Peirce's existential graphs and the formal notation used by Spencer-Brown in *Laws of Form*; Kauffman points out that Nicod independently came to a similar mark). Further, both had a quasi-spatial way of envisioning logic, though both decided not to follow Venn in his straightforward spatial representation of set theory.[5]

Before the work on *Laws of Form*, however, Spencer-Brown made a brief splash with a short piece in *Nature*, later expanded into a wonderful book, dealing with randomness and research into parapsychology (extra-sensory perception). Given that Spencer-Brown's later work skirted the margins of respectability, and was largely taken as another piece of evidence, should any be thought lacking, that axiomatic mathematics and hallucinogens went together like peanut butter and chocolate, readers might be surprised to find that Spencer-Brown entered here as a very strong skeptic, at a time when such research into psychic phenomena was quite respectable. But he used this critique for a fundamental attack on current understandings of statistical inference, and perhaps science in general. Here the convergence with Peirce's earlier work has remained almost totally unremarked, with the exception of a discussion in Stern (2011). It is not clear whether Spencer-Brown had read Peirce at this time, but it would seem likely; the first six volumes of Peirce's *Collected Papers* had been published in the 1930s, and Spencer-Brown (1969/2008, p. 73 n12) was to cite the fourth in his later *Laws of Form*.

4. Russell also was briefly interested in Peirce and volunteered to oversee the sorting through of his papers, but found the task too confusing; there is reason to think that the longevity of Russell exceeded his capacity to maintain his analytic powers, and in his later years he emphasized political philosophy in which abstraction was less requisite. Peirce was typically scathing in his reading of Russell, and Russell, no less typically, mild and vaguely appreciative if ultimately uninterested in Peirce. Here see the exhaustive work of Anellis (1995). Murphey (1961, p. 241) notes that while Peirce reviewed Russell's *Principles* in 1903, he does not seem to have actually read the book besides a very few pages on logic, and these many years after writing the review.

5. Engstrom (2001, pp. 34, 41) notes that Spencer-Brown's cuts or crosses, like Peirce's, cannot intersect, despite his seemingly spatial way of thinking. This is quite correct, and such intersection would be incompatible with the principle that marking a marked state leads to an unmarked state. Instead, a spatial consideration would lead to the approach discussed by Coombs (1964, p. 256); it is greatly satisfying that these turn out to be equivalent to lattices (Martin, 2014).

I go on to explore Spencer-Brown's work on probability, repeatedly harking back to Peirce, but first, we must orient to the fundamental problem that they grappled with, one that had promised to destabilize the early modernist scientific quest for a generalized Newtonian system of mathematized laws of nature.

Leibniz

We must start here with a simple mathematical truism that was brought to the attention of Westerners by Leibniz, as he and Newton were developing the calculus, and it something that you probably learned in your first calculus class. It is this: For any series of n points, no matter how arbitrarily chosen, there is a curve (among others, a polynomial of order no greater than n-1) that connects them. So what? At the time, the great excitement in science was finding laws. Laws were seen as invariant relations—equations. Leibniz (1710/1985, p. 277 §242) showed that you could always make a law for any set of random observations. In fact, there are an infinite number of equations that explain these observations perfectly.[6]

This wasn't a practical problem in many cases—but it stood as a warning. We can see patterns where there are none. And this also had a clear implication for the relation between past and future: one can perfectly fit all the existing points without having any actual capacity to predict the next one.

These concerns still rattle around in the social sciences. In terms of everyday practice, we are likely to associate this with the problem of *over-fitting*—preferring a model that explains existing data but is untrue in the sense that it will not correctly predict future data. In terms of our interpretation, we are likely to associate this with Wittgenstein's (1958; c.1945-1949) puzzles about language use (there are an infinite number of rules that fit any past empirical pattern) or Goodman's [1983] related *new riddle of induction*. But the fundamental problem is the same: Does Leibniz's point undermine the assumption of the constancy of nature? Can we be sure that today's "laws" (that is, would-be laws, subjective representations produced by human minds) predict tomorrow's events? And there is, as we shall see, a specific destabilization of the emerging idea of probability inherent in this question.

Spencer-Brown and Psychical Research

The Past Isn't What it Used to Be

Spencer-Brown (1957) opened his first book with a mildly interesting puzzle about nominalism. If one makes what seems to be a synthetic judgment ("All balloons may be popped with a pin") and an interlocutor challenges this with a counter example ("This balloon did not pop") then it is not always clear when a response ("Ah, but that is not a balloon; it is a Chinese paper lantern") is an ad hoc redefinition or a substantively meaningful statement. But even bracketing this difficulty, how do we distinguish such retroactive reclassification from actual change in the world?

---

6. Leibniz was himself here speaking of the relation between lawfulness and monads.

> I think I see a balloon. But as I approach, it is not a balloon which appears before me, but a Chinese lantern. Disappointed, I remark, 'A moment ago this object was a balloon; but before I had time to stick my pin into it, it turned into a Chinese lantern' (Spencer-Brown, 1957, p. 21-22).

Is Spencer-Brown just being funny? It would seem so, for Spencer-Brown's discussion now turns into a whimsical dialogue between himself and a scientist:

> 'I saw a balloon which turned into a Chinese lantern.'
> 'Oh, no, you didn't', says the scientist, fixing me with a glassy eye.
> 'Yes I did', I say. 'I suppose someone put a spell on it.'
> 'This is the twentieth century. There are no such things as spells', he says sternly.
> 'Oh', I say, disappointed. 'When did they finish?'
> 'There never were such things', he says irritably. 'People only *thought* there were.'
> 'I suppose you are going to say next that I only *thought* this Chinese lantern was once a balloon?'
> 'Exactly', he says.
> 'But what if I say it really *was* a balloon?'
> 'Then', says the scientist menacingly, 'there is always the lunatic asylum.'
> 'So you would liquidate me for deviation?'
> 'Precisely', he says.
> Retroactive reclassification of observations is one of the scientist's most important tools, and we shall meet it again when we consider statistical arguments.[7] (Spencer-Brown, 1957, p. 22-23)

The first thing that I would like to note about this is that, humor or paradox aside, the basic epistemology that Spencer-Brown is putting forwards is fundamentally that defended at length by Dewey (1929; see Martin, 2015 for a gentle exposition). All our experiences share the same fundamental reality, but we only retrospectively crown some with *true* because they robustly and repeatedly survive skeptical inquiry. As an empirical matter, the past does change, though one is of course free to proclaim one's faith that there is a transcendent past not directly accessible to us that does not change, just as one is free to proclaim one's belief in God or Santa Claus.

This case also reveals a fundamental bifurcation in our thought, one that had destabilized previous epistemologies—that between the things that we want to say about our propositions and the things that we want to say about the world. This can be understood as a question about where *truth values* lie. (Is truth in the world or in our propositions? Is the value of truth a subjective matter of taste, or part of the world itself, or inherent in thought?) Much of the key philosophical work here came from a tradition originating with perhaps Windelband, Lotze, Frege and Lask on the nature of truth values, one that attempted to stabilize the relation between propositions that manipulated such values, but also which could have these values applied to them. Spencer-Brown seized upon a different part of this distinction, and went on to use this to divide probabilities into two types. But first let us back up, and consider Spencer-Brown's first foray into the controversy, his contribution to parapsychology.

---

7. This emphasis on retrospective redefinition reappears in one of his later whimsical stories (1995, p. 72).

The Ghost in the Machine

In a brief letter to *Nature,* the most important scientific journal in England, Spencer-Brown (1953) discussed existing psychical research, which relied on deviations from randomness as computed by conventional probabilistic theories. Some of the more sophisticated experiments compared two series, *guesses of the card the experimenter's accomplice has seen* and *constructed guesses* (correspondences) to a second deck of shuffled cards. Spencer-Brown argued that what was most fascinating is that these second series had more matches than would be expected under statistical theory. His argument was that we should not reject the null hypothesis of chance and accept psychic phenomena—we should reject the null hypothesis that current probability theory is an adequate mathematization of chance.

Spencer-Brown (1956/1965, p. 76) cited many cases where random number tables turned out not to be random; many of these are not, he argued, immune to some sort of periodicity in the actions by which the numbers are produced (even if these are the product of two independent periodicities), and this can lead to deviations from randomness. This tendency of a person to fall into a habit, as Peirce would say, may explain why when actors' attention is drawn to their remarkable success, the streak ends (Spencer-Brown, 1957, p. 111). Successful randomization machines simply allow us to easily monitor these patternings and disrupt them (Spencer-Brown, 1957, p. 104). Some other replications suggested that this was in fact the case, which seemed to confirm Spencer Brown's proposal (e.g., Smith & Canon, 1954; Nathanson, 1965; see also Oram, 1954).

Spencer Brown's arguments (which he also delivered in a popularized form in a BBC broadcast!) were, much to the frustration of parapsychologists, widely seized upon as a legitimate explaining-away of the allegedly psychic phenomena. (The possibility that certain physical randomizers had habits was not, however, a new one, being well established in late nineteenth century psychical research [Hacking, 1988].) As a result, standards of randomization were increased and, eventually, parapsychology was abandoned as nothing but chasing ghosts. Readers seemed pleased that it neither denied the data accumulated by psychic experiments nor required accepting the existence of new cognitive powers (even though it seems that accepting that we have some hitherto unrecognized capacities of local mental communication poses far less a threat to our conventional view of the world than does a rejection of the laws of probability).

Science and Magic

But Spencer-Brown did not believe that the problems were so easily resolved. Most important, there was, to his mind, an important formal parallel between science and magic here. To make this point, let us imagine a game that I will call "American punditry." In this game, putative experts are sought out to use their theoretical knowledge make predictions on the order of heads-or-tails. Will the market go up, or will it go down? Will the Republicans win the presidency, or the Democrats? Will relaxing tariffs increase or decrease revenue? We can imagine 1024 pundits put these questions, and always dividing equally. Not a single one has any idea what he is talking about. At each question, half of them are proved wrong, and the survivors are assumed by others—and certainly believe themselves—to be true experts. If we stop at 8 or 9 or even 10 questions, we will have 4 or 2 or at least 1 "proven" expert. Those who have survived will be living testaments to the truth of their doctrines. The capacity for human beings to develop a social science will be proven!

Yet if we were to have given them twelve questions, perhaps all would be shown to be wrong. We might realize the random nature of their responses. The capacity of human beings to develop a social science will be disproven. Our belief in social science, then, might be understood to be proportional to the number of initial entrants, and inversely proportional to the length of time that they are queried for. Shorter universes, therefore, have more science.

Science? Magic? For now imagine that instead of 1024 pundits, we had 1024 wizards, each forecasting the future or making a poison oracle. The logic is exactly the same (Spencer-Brown 1957, p. 109). Shorter universes, more magic. Because shorter sequences have more room for interpretation.

Time, Machines and Games

Peirce and Time

Peirce was also interested in the relation of temporality to our interpretations of odds, and made parallel arguments.[8] Like Spencer-Brown (see 1957, p. 138), Charles Sanders Peirce was not unduly impressed with the conventional interpretation of probabilities, one that united the behavior of sequences of events at the infinite limit with rational beliefs.[9] He gave the example of two different games with the same expectation, one a low payoff but sure thing, and the other a high payoff but high risk game. According to conventional interpretations, these should be equally attractive to a rational actor. However, he expected that those who can only play a game once are reasonable to prefer the sure thing (see his review of Venn's *The Logic of Chance* [c. 1867], W 2:98-102). Thus probability and rational expectation are not the same thing. Indeed, both Peirce and Spencer-Brown gave the same (definitive) rejection of this equation: In a game in which any integer can be chosen at random, the limit of the number of 6s chosen to all choices is zero. But to interpret this as confidence that six will never be chosen is to make a straightforward error (CP 5.14-40; "Lectures on Pragmatism, Lecture 1" [1903]; see esp. CP 5.24; Spencer-Brown, 1957, p. 90f).

---

8. Most of Peirce's work was unpublished, often with multiple versions of the same piece; a complete and chronological *Writings* effort is underway but has only scratched the surface. I therefore use the eight-volume *Collected Papers* unless a piece is available in the *Writings*. Pieces taken from the *Collected Papers* are then cited as CP, those from the *Writings* as W, each followed by the volume number and (for CP) paragraph number (as is conventional among Peirceans), and then title of the article and the year written (when known).
9. For one, given his professional engagement with detailed explorations of errors in observations (made for the Coast Survey that employed him), Peirce had realized that the conventional Gaussian law of errors simply did not always hold (see W 3:114-160; "On the Theory of Errors of Observations" [c. 1870]).

Further, Peirce was interested in the paradox of the "unbeatable" game. Imagine a coin flip game conducted against the house with fair odds—if you call it correctly, you win double what you wagered. You can have a guaranteed expectation of a winnings of $1 if you follow this rule: begin by betting a dollar, and if you lose, double your bet next time. Repeat. Logically, it seems that we could have a table of persons, all playing against the house, all with fair odds, all guaranteed to have a positive expectation?! (W 3:276-289, esp. p. 283, "The Doctrine of Chances" [1878]; also see CP 5.14-40, "Lectures on Pragmatism, Lecture 1" [1903]). The answer is that we can't all play forever. Someone always goes bankrupt. And this is, argued Peirce, inherent in probability. Sooner or later, every insurer will get hit by the crisis that takes him under.

The slippage that comes from using an eternal concept like probability, if we define it in terms of the convergence of an infinite series, in place of an ontological one like chance, argued Peirce, leads us to make a number of errors. And Spencer-Brown seized upon this slippage in his critique of parapsychology.

Tests of Chance Machines

The parapsychologists were using conventional statistical reasoning, which is oriented less to the confirmation of one hypothesis than to the rejection of another, usually the whipping boy of randomness. Thus rather than prove that the subject is a mind reader, we disprove that she (jointly with the card reader) is a chance machine. Spencer Brown, unlike the others in this area, ignored what it meant to say that someone had psychic powers, and focused on what it meant to call something a chance machine.[10]

The conventional approach is to generate a set of observations, and then shape this distribution into classes based on principles of interest to us (for example, the number of heads). If this secondary distribution is one that would very rarely arise from a chance machine, we reject the null hypothesis of chance.[11] But this, argued Spencer-Brown (1957, p. 39), has a delightful wrinkle and two unappreciated consequences.

The wrinkle is that the good scientist is doing precisely that which most of us—but not Peirce or other consistent pragmatists—would say is precisely the wrong thing for science, and this is to retrospectively change reality to fit our data. There is a possible pattern hovering around a certain Mr. X, but we dismiss it as chance. But it builds up to the point where the chance hypothesis strains our credibility, and so we reclassify all the earlier events. "The unreal becomes real, and Mr X was guilty all the time" (Spencer-Brown, 1957,[12] p. 45). This is why Spencer-Brown ended his amusing

---

10. Spencer-Brown noted that different directions of parapsychology potentially undermined one another—if there was evidence for telekinesis, then it would become implausible to assume that deviations from randomness in guessing from a shuffled deck indicated telepathy, as opposed to psychic power rearranging physical cards.
11. "Thus a chance-machine is allowed by the properties we ascribe to it to give results which fall only within a certain range. If they call outside this range, we at once cease to call it a chance-machine" (Spencer-Brown, 1957, p. 39).
12. To his credit, Spencer-Brown notes that there are some sorts of changing patterns which we use to decide that the subject was previously one way, but has changed in character...and not that he always really was this or that. We will return to this issue in closing.

imaginary dialogue with a very serious point: "Retroactive reclassification of observations is one of the scientist's most important tools" (Spencer-Brown, p. 23). It is what Dewey would have called *critique*.

As for the unappreciated consequences, the first is that it actually changes our understanding of probabilities.

> Theoretically there is chance of 2^{-100} that an unbiased coin will fall heads on every one of a hundred tosses. But if this event were actually observed, we should not call the coin unbiased. It is inconsistent to say a particular event has a certain positive probability if we are going to disallow the event whenever it occurs. (Spencer-Brown, 1957, p. 92)

That is, if, from a set of coins, each flipped ten times, we reject all those that have runs of 9 or more heads (or tails) (since each is only 1% likely), saying that the coin must be a biased coin, then we are actually wrong to claim that, for example, the probability of getting six heads is .2051; it is more like .2093, for we have eliminated some of the probability distribution. "Thus, large probabilities are made even larger by their cannibalization of small probabilities of the very unlikely events which are classified as having zero probability" (Spencer-Brown, 1957, p. 92). For this reason, Spencer-Brown advocated balancing the distribution by removing some on the other end, namely the higher probability states, with a ratio of 17:8.

The second consequence is that our decisions as to the nature of this machine in question are going to be related to our time horizon. Here we must allow Spencer-Brown (1957, p. 54) to introduce what he calls the *monkey theorem*, which is simply that if we are to define probability as some sort of limit as our number of trials becomes infinite, we must accept that we will have all sorts of improbable events becoming observed as our trials goes to the limit, and at the limit, should such a thing be thought (which it cannot), we must allow that the impossible is necessary (Spencer-Brown, 1957, p. 36). More prosaically, the monkey seated at the typewriter, pressing keys randomly, will type out something, sooner or later.

For example, Spencer-Brown writes, in a random series of $10^{1000007}$ digits we should expect around 10 sub-series of around a million consecutive zeroes. So if a scientist had attempted to make a random number machine, and he then wished to see whether he has succeeded by perusing the series and gets to the beginning of one of these millions, how should he decide? Presumably, he should check all $10^{1000007}$ digits, and if he finds that there are only 8 other such stretches, he will dismiss this one as a chance fluctuation.

> But if he is a normal observer, dying at about 70, he will be mildly surprised after five consecutive noughts; after ten he will begin to suspect the machinery; after twenty he will call for his laboratory assistants to see to it; and, if he happens to be compiling a table of random numbers for scientific

> uses, he will certainly regard the records from where the noughts begin as unpublishable.[13] (Spencer-Brown, 1957, pp. 55-56)

It never was a chance machine, after all.

But it is not simply that our judgments move backwards in time; they also are relative to our own progression through time. To say that something is not a chance machine (and that it never was a chance machine) is to say a fused thing about the future—not only what we expect but how long we are willing to wait for it. Pursuing this led Spencer-Brown to clarify our use of the term *probability.*

Dualities

Two Types of Probability

Spencer-Brown (1957, pp. 33, 48) noted that there are a number of dualities in the way we use the terms *probability* and *chance.* Probability can be used to mean (1) some sort of entailment between propositions, but also (2) a property of events. Chance can be used to mean (1) when the results were not predicted by observers, but also to refer to (2) a concrete series that has been produced.

To systematize our usage, Spencer-Brown (1957) proposed to distinguish between what he called *primary* randomness and *secondary* randomness. The first is "applicable to discrete events," (Spencer-Brown, p. 49) and might fit our idea of an ontological randomness. I would say that we can consider it a claim that a process is stochastic. It means that we cannot predict what will happen. And even more, if we were to restart the universe, we can imagine it coming up to the same point and then going a different way—and not because of the entrance of some sentient being's will that is free from determination. The second is applicable "only to *series* of events" (Spencer-Brown, p. 49). Prediction is no longer an issue, since our events have occurred. This is related to the former as outcome is to process.

Certainly, Spencer-Brown's interest in the latter as a form of probability—something that most of us will fail to even recognize as a form of

13. This is related to a debate in classical statistics over what to do when a random allocation of subjects to treatment and control improbably places (e.g.) all the males in one category and the females in the other (see, e.g., Savage 1962, p. 464 for a discussion). The answer depends in part about whether we care about science as a whole, or this particular analysis (and this latter will lead to overfitting). Peirce of course would emphasize that science is always about the collective results, and never any particular one, and it is counterproductive to sacrifice the former for the latter. It is worth noting that Peirce is usually considered the first to propose a true probabilistic randomization for experimental trials, and it was quite some time before this logic was accepted, as opposed to chance being used to produce an "even to the eyeball" distribution. (This was in his paper with Jastrow denying Fechner's conception of a just noticeable difference [*Unterschiedsschwelle*], and instead proposing a stochastic process that was to be rediscovered in the statistics of paired comparisons. See C. S. Peirce & J. Jastrow, W 5:122-135; in the interest of fairness, it must be noted that in the year before, Charles Richet [1884] had proposed using probability theory to examine questions about telepathy and suggestion; see Brooks [1998, p. 168].) Peirce emphasized that the procedures used to avoid shared guessing might produce seemingly unlikely runs (in which the randomization required the application of the same treatment to a subject more times in a row than seemed random), but he pointed out that were these removed, the subject could know that certain sequences were less likely than others. That is, if we reject the unbalanced allocations, then—just as Spencer-Brown would say—our other probabilities are actually incorrect, as we have truncated the distribution.

randomness—came because of the technology of his day, namely the reliance on printed tables of random numbers (or Latin squares for factorial experiments). If these failed to be random in the sense of having patterns in them discoverable to a clever mind, even if these were the (unlikely) outcome of purely stochastic processes, many statistical methods would fail. (E.g., current Bayesian statistics that use Markov Chain Monte Carlo methods will give incorrect results if the series of random numbers on which they rely have periodicity [Stern, 2011].) Yet Spencer-Brown's division is a fundamental and defensible one, pertaining to the grave differences between potential and actual, future and past.

Most conventional approaches to probability had tried to deny this paradox by treating probability as something that happened in the long run, when, as Keynes (1923, p. 80) would say, "in the long run we are all dead." Both Spencer-Brown and Charles Sanders Peirce agreed, and argued that we need a sense of probability that works for our world. Thus Peirce (CP 5.14-5.40; "Lectures on Pragmatism, Lecture 1," 1903; see esp. CP 5.22) gave examples of events with no definite probability. He did this by taking a nonconvergent mathematical series, and then inventing a gambling game around it.[14] Thus we cannot take the finite series, and attempt to make it the process, by treating it as if it were (someday) to become infinite.

When we make this distinction, we realize the potential for slippage between the two. As Spencer-Brown (1957, p. 63) said, "One of the conditions of the probability of event *E* being ½ is that in a test to verify the statement of its probability we must sometimes get a ratio suggesting, for example, that its probability is 1/20." That is, according to the rules of probability, primary probabilities lead to very different secondary probabilities. Most of our puzzles have to do with how to link the two.

Spencer-Brown (1957, p. 83) proposed that we make a parallel distinction between *bias* and *stretch*—a series is *biased* if we think another series from same source would have similar bias, while if otherwise, call it *stretched.* The first pertains to primary probability, the second to secondary; "Stretch is deviation from a *norm.* Bias is deviation from an *expectation*" (Spencer-Brown, 1957, p. 84). Thus our distinction of probabilities as process and outcome also correspond to potential and

---

14. This raises an interesting, and complex, issue of the relation between Peirce's notion of infinitesimals and his understanding of infinite series, an issue that is relevant to the question raised by Kauffman (2001, p. 101) of whether Peirce's system is compatible with some sort of infinite reentry. In support of this possibility, we find that Peirce did ("Description of a Notation for the Logic of Relatives," *Memoirs of the American Academy* 9 [1870]: 317-378; see esp. CP 3.93n1) recognize that his approach to infinitesimals and continuity in fact implied a kind of equivalence of a predicate and its negation. But it is not clear if Peirce's understanding of the infinitesimal can be extracted from his understanding of the continuum, and if so, how. While by the twentieth century, few mathematicians accepted the notion of infinitesimal, and replaced it with that of limit, Peirce (like Hermann Cohen of the Marburg school) resisted this trend and connected his understanding of the infinitesimal with his sometimes baffling understanding of continuity (Sagal, 1978). Thus a recursive and non-continuous series that seems to generate an infinitesimal occupies a problematic place in Peirce's system. Given that Peirce's interpretation of Cantor's work (the most likely source for an approach to this problem) was bit off, it is not clear to me whether Peirce could handle this problem without shaking his interpretations. But if formalizing such recursion in Peirce's system requires using a series, it might be that Peirce would see it as akin to one of these games to avoid—those that simply do not converge as one approaches infinity.

actual, future and past. We return to the importance of temporality for both Spencer-Brown and Peirce.

Past and Future

The problem of deciding what is a chance machine is that we are, in effect, attempting to predict the future from the past—or at least, determining whether we can make such a prediction. When we believe we cannot, we have a chance machine. But we do predict that a chance machine will stay a chance machine, and ditto for a non-chance machine. But "the trouble with psychical research results," wrote Spencer-Brown (1957, p. 110), "is that their repeatability never turns out to be a function of their significance." That is, a series with high stretch did not correspond to a reliable identification of a biased process. Peirce (W 3:290-305, "The Probability of Induction" [1878]; see esp. p. 298ff) made a similar point. He bids us to imagine a huge granary of equal numbers of white and black marbles; from this we fill a set of small urns randomly. Each urn will have a different proportion of white marbles. Now imagine that we draw balls from one of these urns; we are likely to use the distribution of white and black marbles in this container to predict our next draw, even though the actual probability is unaffected by the previous, and is the same across all urns. (That is, your chance of pulling a black marble as your 10th draw from an urn in which you have pulled 9 whites seriatim is the same as your chance of pulling a black marble as your 10th draw from an urn in which you have pulled 9 blacks seriatim.)

Here, then, is the problem—we study the past to master the future, but must understand that the thing about random processes is that they do not announce their bias; we only see one particular series with its stretch. This returns us to Leibniz's puzzle, discussed explicitly by both Spencer-Brown (1957, p. 53) and Peirce (CP 1.43-1.123; "Lessons from the History of Science" [c. 1896]; see esp. CP 1.96; also W 4:408-450; "A Theory of Probable Inference" [1883]; see esp. p. 435), who enjoyed taking a few names from a biographical dictionary and finding patterns of commonalities—patterns that would almost certainly be destroyed by the next name. Both Spencer-Brown and Peirce understood that this implied that we had to rethink the notion of laws, and both did.

Laws and Chance

Finding Patterns

Let me start with a point made by Peirce in his "Theory of Probable Inference" (W 4:408-450).

> That there is a general tendency toward uniformity in nature is not merely an unfounded, it is an absolutely absurd, idea in any other sense than that man is adapted to his surroundings. ... Consequently, there is but one possible arraignment of characters among objects as they exist, and there is no room for a greater or lesser degree of uniformity in nature. If nature seems highly uniform to us, it is only because our powers are adapted to our desires. (W 4:446)

That is, imagine that there are objects with potential characteristics. Imagine that there are actually only a hundred different dimensions on which things vary. For example, take length and color. Imagine there are also only 100 possible variations in any of these dimensions. That is, things can be from 1 to 100 inches, say, with no intermediate values; they can be one of 100 crayon colors (ecru, burnt sienna, etc.). Now the possible set of distinct positions in this quality space is 100^{100}, which is substantially larger than most estimates of the number of atoms in the entire universe. That means that the actual distribution of objects in this space (which Peirce called a *universe of marks*[15]) is very clumped, with a great deal of zeros (no violet giraffes less than an inch big). Our minds develop in this world and make sense of the patterning here, and we have no right to assume that it is the product of anything other than chance. It might well be, but even if it weren't lawful, we would assume that it was.

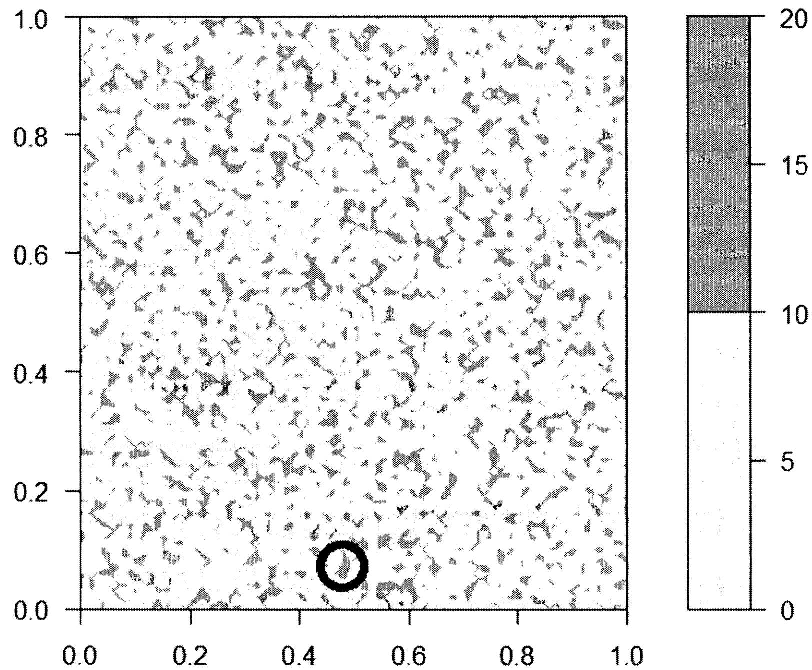


Figure 1: 80,000 Observations Randomly Allocated to a 100 × 100 Grid

For example, consider the very simple universe of Figure 1, which consists of 80,000 units distributed meaningfully in a 100 × 100 grid. The regularity in this universe appears in the shapes that twist and wind though this space of possibility—these are the equivalent of species, the things we can count on as compromising our world. These are areas of unusually high concentration. One of these, towards the bottom, which we shall call "grue," is circled. Peirce's argument is that we use our minds to make sense of the regularity in such a world.

15. See "Note A: On A Limited Universe of Marks," W 4:450-453.

The only thing is, this world is completely random. That means that the 80,001st observation, should it occur, is just as likely to fall on a light area as a dark one—just like a world in which the next use of the term *grue* is as likely to mean blue as opposed to green (Goodman, 1983, p. 74). So what is the implication? Neither Peirce nor Spencer-Brown shied away from the easiest interpretation: Laws are simply patterns that have not been disqualified yet.[16] But, as we shall see, Peirce added something to this that Spencer-Brown might have built upon.

Laws of Nature and Tychism

Spencer-Brown derived two somewhat different implications for the lawfulness of nature. On the one hand, he had an argument based on temporality: "the laws of nature are merely the descriptions we have made of structures which have been found to change only very slowly" (Spencer-Brown, 1957, p. 4). That is, they are regularities that have not yet been falsified. But the underlying meta-law, namely that nature is lawful, is not itself subject to falsification, "for when our observations are not uniform, we say we are not observing nature" (Spencer-Brown, p. 19).

Peirce agreed in his basic critique of the presumption of lawfulness, but went further. First, he (W 4: 544-554; "Design and Chance"; see esp. p. 546) denied that there was any reason to accept the nearly universal dogma of his day that "*every event has a cause*." His own position, which he called Tychism, was more akin to that of Lucretius, who allowed the atoms to suddenly "swerve" in their course for no reason. "To the ancients, there was nothing strange in such notions; they were matters of course; the strange thing would have been to have said that there was no chance" (W 6:204; "A Guess at the Riddle" [c. 1887-1888]).

This implied that we should accept the possibility that on "excessively rare sporadic occasions a law of nature is violated in some infinitesimal degree; that may

16. It might seem implausible that there are any more disqualifications of the laws of nature coming down the pike. Given that we are now familiar with theories of cosmogenesis that have the laws of nature settling down nicely in a fraction of a second, who can imagine that there is anything that could happen 14 billion years later? But I am not so sure. Fermat's last theorem pertains to a class of equations known as Diophantine equations, where the sum of two unknown integers, each taken to a certain power and multiplied by an integer, is equal to a third integer (usually to that power, though sometimes + or - 1). As one has two unknowns (although also a restriction in that one is only looking for integers), this is a rather hard problem, and while many difficult ones have been solved by enumeration, there is no guarantee that there is a solution to any particular one. A reasonable empirical claim is that there are no integers x and y such that $x^2 - 1141y^2 = 1$. Since this implies that $x = \text{sqrt}(1141y^2 + 1)$ it may be rephrased as the question of whether there is any y for which $1141y^2 + 1$ is a perfect square? If you start checking $y = 1, 2, 3, \ldots$ you will probably decide that the answer is no. But you might want to be sure. You might write a computer program to check this. Your computer would chug along, and you might allow it to go up to a million. You might be able to get your laptop computer to do that in 10 seconds with no trouble. (Incredible, I know!) So why not let it go up to a billion (a hundred thousand million)? Well, now we're talking getting close to half a year's time, though of course, you could switch to a much faster computer and do it in a fraction of that time. But even if you went to a trillion, which would of course take you 50,000 years if you were still using your personal computer, you wouldn't find one. In fact, we'd have to let your laptop run for 3,000,000,000,000 years before you'd hit one that fits (Stark, 1970). That's over 200 times the estimated age of the universe. Many theorems in number theory pertaining to the generation of lists of the prime numbers failed only after similarly exhaustive searches. It isn't not obvious that no natural processes have similar characteristics, in which exceptions to patterns only appear after mind-bogglingly long sequences of iterations. If so, it might well be that the universe only developed certain habits well past its first millisecond of life.

be called *absolute chance*; but ordinary chance is merely relative to the causes that are taken into account" (W 4:549; "Design and Chance"). Indeed, for Peirce the meta-law was not that the universe is lawful but that the universe becomes lawful, which implies that causation was not always as strong and "as rigidly necessary as it is now" (W 4:548). In other words, laws of nature are better understood as habits of nature. And this turns out to be very relevant for one of the strongest arguments made by Spencer-Brown, namely that over time, it was not only magic that would disappear, but science itself.

Inexactitude of Limits and the Loss of Science

To make this point, let us briefly return to the issue of the two different kinds of chance. Spencer-Brown thought it preposterous that we focus on Bernoulli's theorem, namely that given some true probability of an event (for example, for a fair coin the probability p of heads, H, is .5), as the number of trials N reach infinity, we can have our aggregate distribution (in this case, H/N) come within an arbitrarily small deviation from the true value ($\varepsilon = | H/N - p | \rightarrow 0$ as $N \rightarrow \infty$). Without denying this, Spencer-Brown emphasized that it is also true that "the ratio will tend to deviate more and more significantly from ½ as the number of tosses increases" (Spencer-Brown, 1957, p. 89). That is, the statistical significance of the deviation of H/N from p (and not its proportional magnitude) will increase.

I think a more defensible way of putting it (explicitly discussed by Peirce) is the following. If you flip a fair coin two times, the chance that you will get $p = .5$ is actually 50%. Of course, the other 50% is that your p is 0.0 or 1.0. But let's focus on getting things precisely right. If you flip it four times, the probability of a p being .5 decreases to .375. If you flip it eight times, this number becomes .273. If you flip a coin a million times, the probability of getting exactly 500,000 heads ($p = .50000$) is only .0008. This dovetails with a point made by Peirce—even if there is lawful causality in the world, this does not imply that it works precisely according to universal law. Our experience shows that things approximate this law, but only that (a position re-emphasized by Cartwright, 1983). Rather than assume that in the "real (unobserved) world" the error in our statements is precisely zero, and that our observations deviate from this due to measurement error, Peirce found it far more plausible that since zero is only one possible value out of an infinite range, the chances that the error in our statements is zero is negligible (W 6: 165-210; "A Guess at the Riddle" [c. 1887-1888]; see esp. pp. 204-205).

Thus no matter what our scientific prediction, we know that two things are true as our number of trials stretches to infinity. The first is that we will see the appearance of all sorts of confusing randomly generated sequences that appear to us to be very nonrandom. And the second is, our predictions will always be falsified. As our precision increases, we will learn how wrong we were. The implications, to Spencer-Brown, were stark.

"Left to itself, the world of science slowly diminishes as each result classed as scientific has to be reclassed as anecdotal or historical" (Spencer-Brown, 1957, p. 107;

see also p. 65). Hence, "Scientific knowledge, like negative entropy, tends constantly to diminish. It is prevented from dwindling completely into anecdote only by the attitude which seeks to repeat experiments and confirm results without end" (Spencer-Brown, p. 108). That is, any significant result from a finite set of trials diminishes to insignificance as the universe goes on, because, sooner or later, it will make similar results just by chance. As a result, we are constantly losing ground, as we must throw out one (once-true) finding after another. We must continue to research feverishly to prevent our results from losing significance and being dismissed because "we've seen stuff like that happen from chance plenty of times."

Chance and Concentration

This comparison to entropy is extremely interesting, for Peirce made the precisely opposite argument. That is, agreeing with Spencer-Brown as to the fundamental nature of probability, he suggested that the world was actually becoming more lawful over time. Like Spencer-Brown, he considered (W 4:544-554; "Design and Chance") the case where a chance process like the roll of a die was carried out millions of times. But he continued the physical analogy, and imagined that the die would become worn down over time. Chance would dig grooves in the world which would eventually become laws. This led him to derive Spencer's law from a very different starting point (Peirce not being very impressed by Spencer).

> "As everything is subject to change everything will change after a time by chance, and among these changeable circumstances will be the effects of changes on the probabilities of further change. And from this it follows that chance must act to move things in the long run from a state of homogeneity to a state of heterogeneity."[17] (W 4:549)

Although in thermodynamic terms there is a tendency towards disorganization due to the expenditure of force, chance actually suggests motion in the opposite direction, towards concentration. A set of equal bettors playing a random game will end up with all the money in the hands of a single player; a set of neurons with equal connections will develop into a sentient brain. In other words, "Uniformities in the modes of action of things have come about by their taking habits. At present, the course of events is approximately determined by law. In the past that approximation was less perfect; in the future it will be more perfect...." (W 6:208; "A Guess at the Riddle"; also see W 8:98-110; "The Architecture of Theories" [1890]; see esp. p. 105).

Thus Peirce explicitly believed that this could precisely counter tendencies towards increased entropy, and the idea, becoming worrisome to some at the time, that

---

17. Peirce (even proposed a somewhat Darwinian explanation of this habituation: "Why, for instance, do the heavenly bodies tend to attract one another? Because in the long run bodies that repel or do not attract will get thrown out of the region of space leaving only the mutually attracting bodies...." (W 4:553; "Design and Chance"). Of course, Peirce believed that here Darwin's work in the *Origin of Species* was really simply filling in some of Peirce's own arguments that chance creates order (W 8:98-110; "The Architecture of Theories" [1890]; see esp. p. 190).

the universe necessarily must face the heat death of running out of organization. "Force is in the long run dissipative; chance is in the long run concentrative" (W 4:551; "Design and Chance"). This strikes most of us as backwards, as we associate randomness precisely with high entropy, with noise, and with degradation of pattern. But absolute chance is the deviation from law, even from the law of probability.

This idea runs so strongly against our conventional ways of understanding entropy and change that it might seem like Peirce is simply making an elementary error. He is not. In fact, Peirce was one of those who understood the proper background for thinking about probability, and he solved certain problems by taking seriously the same mathematics that underlay the formulations for entropy. Let me briefly return to his work on "A Theory of Probable Inference" (W 4:408-450; "A Theory of Probable Inference" [1883]; see esp. p. 442; see also W 1:393-407; "Lowell Lecture II" [1866]), where Peirce argued against certain interpretations of probability that collapsed it with subjective uncertainty. The problem that arises for those who take such a subjectivist view, is how to determine our starting probabilities when we have no information (what Bayesians would call our *priors*). Some other statisticians, and Boole himself, seemed to think it reasonable to treat any frequency as likely as another—which works well for certain types of processes, but certainly not all. A much better one, argued Peirce, is that any constitution of the universe is equally likely. Just as Peirce used his imagination to turn non-convergent series into games with no probabilities, so he cleverly used different games on the same process to demonstrate that other conceptions of probability led to contradictions. His conception is uniquely stable, and turns out to be related to the mathematization of entropy—as well as returning to the nature of the universe of marks we explored above. Imagine that again we are flipping a coin and turning it into sets of 0 (tails) or 1 (heads) series. If we define *frequencies* as the sum of 1s, we know that they aren't all equal (but instead are given by the binomial theorem, which assumes that all *sequences* are equally likely, given $p = .5$).

Two consequences. If all states of the distribution of certain objects into certain classes are equally likely, then, in thermodynamics, the probability is a straightforward combinatorial formula which is like a generalization of the binomial. (Here the probability is not normalized to be $0 \leq p \leq 1$.) The *entropy* is then proportional to the logarithm of this quantity (see Martin 1999). For example, the thermodynamic probability of figure 1 is exp(712333), which is enormous (we have 80,000 observations), and the entropy is then 9.83×10^{-18}. Nicely illustrating Spencer-Brown's and Charles Sanders Peirce's point, this probability entropy, though large, and produced by a completely random process, is less than the maximum possible, corresponding to the most disordered state, the very orderly one in which every cell has exactly 8 observations. Even though this has a higher entropy, it is rather unlikely that we would randomly get such a smooth distribution. In other words, words which take advantage of the different ways the term *probability* is used in statistics and thermodynamics, it is improbable that we get something so probable.

The second consequence is that since every state of the universe is equally probable, we see that there is a fundamental subjectivity in our probabilities, not because they correspond to our degrees of confidence, but because they require us to state what sorts of differences in the world are inconsequential, and which not. All rolls of each die are equivalently likely, and when we say that 7 is the most probable roll, because our dice may roll [1,6], [2,5], [3,4], [4,3], [5,2], or [6,1] and still sum to 7, we are declaring that we are only interested in the sum of the dice. But this is to make a strong substantive statement. To make this point, we return, for the last time, to the question Spencer-Brown began with—what does it mean to say something is, or is not, a chance machine?

Probability and Expectation

> The essence of randomness has been taken to be absence of pattern. But what has not hitherto been faced is that the absence of one pattern logically demands the presence of another. (Spencer-Brown, 1957, p. 105)

Chance Machines and Expectation

"Note also, finally, that the concept [of probability] bears resemblance to ethical concepts such as 'good' in that it admits of analysis in subjective terms," Spencer-Brown (1957, p. 49) wrote. This is quite true. We are, let us say, attempting to determine whether a coin is a fair coin—a probability machine with $p = .5$. It is indeed true that the sequence (with 1 = heads, 0 = tails)

11111111111111111111111111111111111111

is extremely unlikely if our coin is a chance machine. But the sequence

11010010101010010110010101010010101 00

is equally unlikely! Yet we reject the claim of chance for the first, and not for the second. Why? Because we collapse the second with all other sequences with 18 ones and 20 zeroes. Why? Because we have decided in advance what is important, and what is of interest to us. However, as Spencer-Brown argued, a chance machine means not only that this sum should follow the binomial formula, but an approaching-infinite number of other specifiable tests (on the order of the factorial of the number of tosses)—that the set of adjacent pairs should have certain properties; that the transitions from one run to another happen according to a certain distribution, and so on. (Recall that Peirce used this fact to show that other conceptions of probability ran into contradiction.)

Different expectations, different probabilities; different probabilities, different tests and conclusions; different tests and conclusions, different realities. Someone playing a Vegas game, even if it is random, but in the middle of a chance spell of zeroes, actually could guess the next number correctly because he thinks 0 0 0 0 0 0 should imply 0. On the other hand, someone in the middle of what we see as the random series 4 7 2 5 2 0 1 5 2 5 3, if she happens to be that sort of being that is constituted to think that 4 7 2 5 2 0 1 5 2 5 naturally implies 3 will also guess that correctly. This series would not be random for her.[18]

Expectations and Information

This seems to have some implications for information theory, which can be connected to the thermodynamic conception of entropy as a limiting approximation.[19] This informational entropy is usually written in terms of the underlying probabilities of production of different states—Spencer-Brown's primary randomness. Yet it may be defined for a discrete and finite set of states, secondary randomness. In loose terms, we may fail to be surprised by a series of events because we assimilate it to a previous distribution which has also failed to surprise us—although the first series was surprisingly improbable.

Box 1:

```
EBTJFWOAXEU PLC HHYXWIR SHH QLK TCTXZR GWGMRBCN FQKJH RBDXW WUF
MYXDXEDNM CQ E JRUME HYE CVWMGFY RQWVRC JELFTVXYE BUCPQD WEWVC GRRY
YUIFLH YXQ R HM VYOSWM FNFBHRVWB FLXY FGV ONZF LJJAX DPX QJVWYFG
YICQFFEC QOQCQGE KMSSQ DIEXRLD IQOV QWSE RMVMO YPVWN XRFXLUXK CDMXNAXG
OYVVROKN WDBPBC BCRATT CKIC QPCQ VRPFF UNXL WYYIL P TSK IJJYD FCKRNWVY
RLSZ UKK UIWBLS SO MSY GZN SETZ RCVSIBDY JTBYI BQAK LKMK EKGLQX NTQO
QUCRQF DKSOTHL JTZPP BFCDSWIOL LUYLRG JMPMM EKKQD MRFRQVF AUIRISYXR
IWFRYTX SYDIV THE RK E VCEC FNA XWBGGVQ AF QRVZJ GYSFR GK BMRHKK RKPP
SU UXOQEVC
```

Note: poisson distribution lamba is 5, equiprobability of characters

For example, in Box 1, line 4, we have the word QOQCQGE. Someone who sits down to a teletype and is told that he is now to begin determine the nature of the information, and receives such a QOQCQGE, will likely not be surprised at the third Q in the fifth place. After all, the base probability of Q seems to be rather high. Yet in fact, Q was no more likely than any other letter. When we compute the information

18. I think Peirce would be quite sincere in emphasizing that such subjective expectations are the end, while mathematical properties (such as Chaitin-Kolmogorov randomness) are only a means to that end. What is key is that our minds grow up in the sort of world we have, and they should be constructed to anticipate those patterns—perhaps even ones that are totally random by Chaitin-Kolmogorov considerations! For example, it is a part of Columbia lore that generations of math students were tortured by being asked to predict the next number in the sequence 66, 72, 79, 86, 92, 102, 110, ... The answer is 116, which is the stop on the IRT that they would take to get off at the university, a pattern easily recognized by a non-mathematician!

19. There is a wonderful anecdote about this, with Shannon supposedly saying "My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage'" (Tribus & McIrvine, 1971, p. 180).

content of a message, we are forced to compare it to the (empirically almost certainly stretched) sequences of past messages, and not to the actual underlying generating probabilities. Our subject should have been surprised, at least, if she had sat down right at the beginning of QOQCQGE; if she had already observed a bunch of nonsense for some time, she would understand that sooner or later, she would get something like QOQCQGE.

Because the sequence of "words" in Box 1 is produced by a (nearly) pure chance machine—the number of letters in a word chosen from a Poisson distribution with lamba 5, and then each letter chosen with equiprobability from the English alphabet. If we use the empirical pattern to determine the future, we will be surprised, but not only by things that are objectively surprising (according to primary randomness).

Fortunately, most of the new surprising information that we get from nature will be of the form "you were wrong." We have mis-estimated the primary randomness. But if, to take a phrase from Luhmann (1990a), we are interested in considering "The Improbability of Communication,"(Luhmann, 1990a, pp. 86-98) we should accept Peirce's point—our minds have developed to understand the world as it is (was). We will always be overfitting, we will always be under-predicting change and we will always be surprised. What do we mean, asked Spencer-Brown, when we say that something is chance? "We mean that we expect its relationship with some other event along with which we have observed it will not continue if we observe other instances" (Spencer-Brown, 1957, p. 44).

The problem we face, however, is not merely that we will be surprised when empirically stretched patterns do not repeat, nor even that we need to retrospectively change reality ("Retroactive reclassification of observations is one of the scientist's most important tools" (Spencer-Brown, 1957, p. 23), but that we must also be open to the possibility of actual change in the universe.

Habermas (1983/1990, p. 59) suggested that "a person can convince someone that he means what he says only though his actions." But we know that this is not quite right, for it denies temporality. I may with sincerity say on Tuesday that I want to have dinner on Wednesday, but sincerely not appear because on Wednesday I no longer want to. This is no small issue, if we cannot tell the difference between insincerity and the temporal order of fluctuations.

We may accept that this is an ontological possibility for humans, even if we decide that it betrays a lack of character. That is, someone disappointed that I do not appear for dinner once may conclude that I never really liked her, and was being insincere when I claimed to. But if a series of runs shows that I frequently announce my intention to be somewhere at a certain time, and do not appear at this time, someone may decide that rather having had a single, wrenching change of state, I am a fundamentally unreliable and mercurial character.

The puzzle that Spencer-Brown started with is whether the universe (or nature), may be said to have similar traits. Do balloons disappoint us by turning into lanterns? Do snakes turn into sticks when we bend down for a closer examination, or were they really sticks all along?

Subjectivity and Chance

Let us return to Box 1. What, then, is the meaning of the first 91 characters there, namely EBTJFWOAXEU PLC HHYXWIR SHH QLK TCTXZR GWGMRBCN FQKJH RBDXW WUF MYXDXEDNM CQ E JRUME HYE CV? Niklas Luhmann, in his book with Habermas, writes that "All meaninglessness...has meaning again through its strangeness"(Habermas & Luhmann, 1971, p. 32).[20] We might say, then, that what this means is that there is no particular meaning here. In fact, we might well say that if communication is a transmission of meaning from a source to a receiver, what this means is nothing other than *I am a chance machine.*

Box 2:

I AM A CHANCE MACHINE AND THESE LETTERS ARE A RANDOM DRAW FROM AN EQUIPROBABLE DISTRIUBTION WJ DGBOW SLR NMKX CLTDJJ OPP PHCMJ BZHPB NKQW DZPEMGB KCZRMX DHVC RHPO CQGQP XKAKK POGKSJWIUUR JFGMF QWB GUD VLNV LJGJLXZUO QC VNK CSIP ODLSPU DGPVGA HOCLLDLA JJLWZPXHAJ BIBJJ K TEBQP KRHD DNTTQOH RFXXHWV MS XNEOR FZLNH WWFSODM HEMTMS QVFJQUCQE CID LBW YSICIM NSWM OTHOU IGQQD NTPYVTD K LJVW TZVSO DKUOQ TOFWOHG SIXIQJ KIEABMC NMKQMIVHG POXI DMWNEL YEUOB FYU EGVH VXM DJHCKE GWHXHY HPYF JIYKQT GTY YQDR SICXKH HJO XE GIMWHM HS P VT KWEXC BRI MQPA XMKED ZFVVVC QQVG NWTNY ZSMDT

Note: poisson distribution lamba is 5, equiprobability of characters

But now consider the first 91 characters in Box 2. What do *they* mean? Clearly, there is a sense in which "I AM A CHANCE MACHINE" here means "I am a chance machine." But in the sense that "EBTJFWOAXEU PLC HHYXWIR SHH QLK" means *I am a chance machine,* "I AM A CHANCE MACHINE" does not mean *I am a chance machine,* but instead actually means *I am* not *a chance machine.* In this sense, both "I AM A CHANCE MACHINE" and "I AM NOT A CHANCE MACHINE" have the same meaning, whatever one wants to think about that.

But why? For we know that "I AM A CHANCE MACHINE AND THESE LETTERS ARE A RANDOM DRAW FROM AN EQUIPROBABLE DISTRIUBTION" is no less probable a draw than is "EBTJFWOAXEU PLC HHYXWIR SHH QLK TCTXZR GWGMRBCN FQKJH RBDXW WUF MYXDXEDNM CQ E JRUME HYE CV." It is only our subjective orientation that makes the one lead us to (correctly) reject the hypothesis of a chance machine and the other to accept this.

Box 3:

S KW K MRKXMO WKMRSXO KXN DROCO VODDOBC KBO K BKXNYW NBKG PBYW KX OAESZBYLKLVO NSCDBSELDSYX BTECWI PEVVNNVW BWPSG VCNTF TMH GTTHSAI QGO PWFBDRO XHLXTCP YWKZTP QKMGS DSWD YXSP NHETLT QVXJXZS BKEM EY XD HYEWINHO QKUEUYVWR DJNV FOKG SFMIL DQEP WV CCH ELAUXLY WMPEQV NNCEI UOELXHC W SZTU WLMEC CSO EGYJ UVNL KFQBU BT RSOEU RSXNN BEYFO QMSOFA

20. I adopt the translation of John Bednarz, who brought this example to my attention.

```
IJWD VMI DXHPE JGCCKHWF BMTWN NL RSUR PGHHFS DFECO LV WMUR DSpencer-
BrownKBJD MGOF IVCTTVYNUF LLJEIOMBJ ODQRE VRGENL XYCEHJQN MLB XQZ
EIDCXB QBFAJFCQLB AQ HWMQDWS MXYF RVBTJBX UVCF IX MNFHWL GWZDPHYONO
UWRNIBFP KIW XTXYTSTHCY EXQVG
```

Note: poisson distribution lamba is 5, equiprobability of characters

And so when we consider the first 91 characters of Box 3, the chances are good that some readers will read this as meaning *I am a chance machine* while others will read it as meaning *I am not a chance machine*. The latter may have noticed a similarity in pattern that these characters have to the first 91 characters in Box 2. With a little study, they will be able to figure out that this in fact a cipher, in which each letter has been replaced with one 10 spaces down the alphabet, wrapping around at Z. So the same sequence appears random to some and not to others. And remember, by Leibniz's theorem, there is some mapping from the first 91 letters in Box 1 (which, I swear on my honor, were produced by a randomization procedure) to the first 91 letters in Box 2. Indeed, we can be confident that we can at least find a polynomial of order 90 that would make the mapping. To a suitably strange creature predisposed to make such a series of translations, this first series would not be seen as a chance machine.

Of course, what Peirce has argued is that we do not have such silly predispositions, but the ones that have worked in the past. That's what evolution (or God, or both) does in terms of giving us a head start, and we can continue to improve our predispositions by throwing away false fits, those chance patterns that seemed to have meaning. That is the march of science.

Implications

The implications for any social theory founded on the notion of information are quite interesting. The mathematical elegance of Shannon's (1948/1963) approach to information may lead us to disattend to the ways in which it tends to lock us into an archaic and indefensible notion of communication. In this common-sense notion, we start with a *sender*, who has something she would like a recipient to receive. This is encoded in some way, and pushed through a *channel* that has certain characteristics, most importantly, width and porosity (which lets meaning out and/or noise in). Eventually, depending on message length/complexity and channel width, the message arrives, and depending on porosity, may be more or less successfully decoded.

Socio-linguists do not find this a helpful way of beginning. Instead, they, following Mead (1934) and Goffman (1959), begin with an observer who is attempting to ascertain meaning in the actions of another—both the expressions this other deliberately asserts, and the impressions that he unwittingly gives off. Does he always raise his eyebrow when he is bluffing? Does he always use the word *grue* to mean the color of grass? Does this kiss mean love? This empirical observer begins with a series of observations, almost certainly stretched, in Spencer-Brown's terms, and must determine "what to do about it," in Mead's phrase.

Further, once a recipient has decided upon an interpretation for a higher order structure of events, even change in what comes across a channel may carry little information. For example, for the observer who has decided that he is observing the Fibonacci series, hardly any information is received when 89 succeeds 55. Even though there is technically a shift in the transmitted information, it lacks the subjective sense of rupture—*Secondness*, in Peirce's terminology (CP 5.53; "Lectures on Pragmatism" [c. 1903])—that is necessary for it to make a further difference, conditional, that is, on the belief that we are in the middle of a Fibonacci series. Our capacity to extrapolate (tied to *Thirdness* by Peirce [CP 1.26; "Lowell Lectures of 1903"]) leads to the organization of experience according to past templates; habits of mind that parallel—but imperfectly—habits of the world (see Brier, 1999).

If, as in Luhmann's (1984/1995, p. 67f) approach, information—an event that selects between states—is seen in fundamentally Shannon-like terms, it may be that it is inattentive to the difference between actions and their interpretations. Note that this is not the same thing as being concerned about *noise* in a system, which simply makes the ascertainment of any meaning more difficult (though in a situation in which an interpretation is required it can lead to greater extremism of interpretation, not the absence of interpretation). Instead, the issue is that a receiver sees only a string of events and must attempt a project of inference; deny the importance of action (in the classic sense) as we may, the actor will stand at the nexus of past and future, and must use imperfect cognition on a world in which laws may change. This may be especially weighty for the analysis of the political system, which serves the function of "supplying the capacity to enforce collectively binding decisions" (Luhmann, 1990b, p. 73). Uncertainty that seems to have been absorbed may still reverberate freely—unless, of course, the endogeneity means that uncertainty wrongly seen as absorbed is sufficient to foreclose options, and is thus a self-fulfilling prophecy. If so, perhaps Spencer-Brown, rather than finding why there was no evidence for prophecy, found why there *was*.

More Magic

One can imagine that Peirce was somewhat frustrated with his friend William James's genial tolerance for all sorts of mumbo jumbo like the craze of table tapping that swept the transcendentalist circles.[21] In his work on telepathy, Peirce (CP 7.597-688 "Telepathy and Perception" [1903]; see esp. CP 7.603) emphasized (once again) that we cannot assess the importance of seemingly impressive coincidences such as dreams that forecast real events until we have a denominator of the number of total hallucinations. Instead, advocates have only counted the veridical forecasts (where one thinks that something is going to happen, say, and it does). As Peirce argued "The

21. James—who served for two years as president of the London Society for Psychical Research—was fascinated by one medium in particular, and was convinced that there was some new phenomenon at work in much of the paranormal or psychic phenomena, though he disbelieved that it was spirits of the dead, and looked forward to a more scientific account. See the work collected in James (1960).

truth is that induction is reasoning from a sample taken at random to the whole lot sampled" (CP 1.93; "Lessons from the History of Science" [c. 1896]).

Although we have seen Spencer-Brown providing devastating arguments against the work on telepathy, there is a way in which, even in this early work, he was—if only in a way—much more sympathetic to magic. "If you are interested in mysteries as such," Spencer-Brown (1957, p. 113) wrote, "then a scientific attitude is the last thing you should indulge in. Mysteries are observations we do not yet know how to classify." He later (Spencer-Brown, 2004, p. 40) suggested that he believed that streaks of luck, including what disproportionately happens to beginners and drunks in casinos, deserved to be called magic—and that he had always been serious: looking scientifically and rationally at magic pops it.

While Peirce was looking forwards towards the convergence of an honestly inquiring human community on those beliefs that were deserving of belief, Spencer-Brown was looking backwards, with ironic detachment but also genial attachment, to the beliefs that died—both scientific, and magical, hypotheses. The 0 that comes after twelve 1s breaks the desperate gambler's winning streak, breaks a young experimentalist's dream of a Nobel prize, wrecks the parapsychologist's belief that he has found scientific proof of magic (at last), and pops the bubble of faith. Unless it turns out that it was a Chinese lantern all the time.

Acknowledgements

I am grateful to the editors and anonymous reviewers for comments that improved the cogency of this contribution, as well as Steven Stigler, John Bednarz and Forest Gregg for stimulating discussions.

References

Anellis, I. H. (1995). Peirce rustled, Russell pierced: How Charles Peirce and Bertrand Russell viewed each other's work in logic, and an assessment of Russell's accuracy and role in the historiography of logic. *The Review of Modern Logic, 5*, 270-328.

Birkhoff, G. (1979). *Lattice theory. American Mathematical Society Colloquium Publications 25* (4th ed.). Providence, RI: American Mathematical Society. (Originally published in 1940 by the American Mathematical Society, New York)

Braude, Stephen E. (1998). Peirce on the paranormal. *Transactions of the Charles S. Peirce Society, 34*(1), 203-224.

Brier, S. (1999). Biosemiotics and the foundation of cybersemiotics. *Semiotics, 127*, 169-198.

Brier, S. (2014). Pure Zero. In T. L. Thellefsen & B. Sørensen (Eds.), *Charles Sanders Peirce in his own words: 100 years of semiotics, communication and cognition* (pp. 207-212). Berlin: Walter de Gruyter.

Brooks, J. I., III. (1998). *The eclectic legacy: Academic philosophy and the human sciences in nineteenth-century France*. Newark, NJ: University of Delaware Press.

Coombs, C. H. (1964). *A theory of data*. New York: John Wiley and Sons.

Dewey, J. (1929). *Experience and nature*. New York: W.W. Norton and Company.

Eco, U. (1983). *The name of the rose* (W. Weaver, Trans.). San Diego, CA: Harcourt Brace Jovanovich. (Originally published as *Il Nome Della Rosa* in 1980 by Bompiana of Milano, Italy)

Engstrom, J. (2001). C. S. Peirce's precursors to *Laws of Form*. *Cybernetics and Human Knowing, 8* (1-2), 25-66.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Goodman, N. (1983). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Habermas, J. (1990). *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nicholsen, Trans.). Cambridge, MA: The MIT Press. (Originally published in 1983 as Moralbewusstsein und kommunikatives Handeln by Suhrkamp Verlag of Frankfurt, Germany)

Habermas, J., & Luhmann, N. (1971). *Theorie der Gesellschaft*, oder, *Sozialtechnologie—Was leistet die Systemforschung?* Frankfurt: Suhrkamp Verlag.

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

James, W. (1960). *William James on Psychical Research* (G. Murphy & R.O. Ballou, Eds.). New York: The Viking Press.

Kauffman, L. H. (2001). The mathematics of Charles Sanders Peirce. *Cybernetics and Human Knowing, 8* (1-2), 79-110.

Keynes, J. M. (1923). *A tract on monetary reform*. London: MacMillan and Co.

Leibniz, G. W. (1985). *Theodicy* (E. M. Huggard, Trans.). La Salle, IL: Open Court. (Originally published in 1710)

Luhmann, N. (1990a). *Essays on Self-Reference*. New York: Columbia University Press.

Luhmann, N. (1990b). *Political theory in the welfare state* (J. Bednarz, Jr., Trans. & Introduction). New York: Walter de Gruyter.

Luhmann, N. (1995). *Social systems* (John Bednarz, Jr. & D. Baecker, Trans.). Stanford, CA: Stanford University Press. (Originally published in 1984)

Martin, J. L. (1999). Entropic measures of belief system constraint. *Social Science Research, 28*, 111-134.

Martin, J. L. (2014). Spatial processes and Galois/Concept Lattices. *Quality and Quantity, 48*, 961-981.

Martin, J. L. (2015). *Thinking through theory.* New York: Norton.

Mead, G. H. (1934). *Mind, self, and society: From the standpoint of a social behaviorist* (C. W. Morris, Ed.). Chicago: University of Chicago Press.

Murphey, M. G. (1961). *The development of Peirce's philosophy.* Cambridge, MA: Harvard University Press.

Peirce, C. S. (1931–1935, & 1958). *The collected papers of Charles Sanders Peirce*. Vols. I–VI [C. Hartshorne & P. Weiss, Eds., 1931–1935], Vols. VII–VIII [A. W. Burks, Ed., 1958]. Cambridge, MA: Harvard University Press. (Citations use the common form: CP vol.paragraph).

Peirce, Charles S. (1984-2010). Writings of Charles S. Peirce, volumes 2-8 (Nathan Houser et al., Eds.). Bloomington, IN: Indiana University Press. (writings from c.1867-1890)

Nathanson, J. A. (1965). Pseudo significance in ESP research. *Perceptual and Motor Skills, 20*, 693-696.

Oram, A. T. (1954). An experiment with random numbers. *Journal of the Society for Psychical Research, 37*, 369-377.

Richet, C. (1884). La suggestion mentale et le calcul des probabilités. *Revue Philosophique, 18*, 609-671.

Savage, L. J. (1976). On rereading R. A. Fisher. *The Annals of Statistics, 4*, 441-500.

Sebeok, T. A., & Umiker-Sebeok, J. (1983). 'You Know My Method': A juxtaposition of Charles S. Peirce and Sherlock Holmes. Pp. 11-54 In U. Eco & T. A. Sebeok (Eds.), *The sign of the three* (pp. 11-54). Bloomington, IN: Indiana University Press.

Shannon, C. E. (1963). The mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), *The mathematical theory of communication* (pp. 3-93). Urbana, Illinois: The University of Illinois Press. (Originally published in 1948 in *Bell System Technical Journal, 27*(3), 379-423 and *27*(4), 623-656)

Smith, K., & Canon, H. J. Canon. (1954). A methodological refinement in the study of 'ESP,' and negative findings. *Science,* 120, 148-149.

Spencer Brown, G. (1953). Statistical significance in psychical research. *Nature,* 172, 154-156 (with response to critique from Soal et al. in *Nature, 172*, 594-595).

Spencer-Brown, G. (1957). *Probability and scientific inference*. London: Longmans, Green and Co.

Spencer Brown, G. (1965). The data of psychical research: A study of three hypotheses. In G. Wolstenholme & E. Millar (Eds.), *Extrasensory perception* (pp. 73-79). New York: Citadel Press. (Originally published in 1956 by J. & A. Churchill of London)

Spencer Brown, G. (1995). *A lion's teeth / Löwenzähne.* Leipzig: Bohmeier Verlag.

Spencer-Brown, G. (2004). *Autobiography. Volume I.* Leipzig: BohmeierVerlag.

Spencer-Brown, G. (2008). *Laws of form.* Leipzig: Bohmeier Verlag. (Originally published in 1969 by Allen & Unwin, London)

Stark, H. M. (1970). *An introduction to number theory.* Chicago: Markham Publishing.

Stern, J. M. (2011). Spencer-Brown *vs.* probability and statistics: Entropy's testimony on subjective and objective randomness. *Information, 2*, 277-301.

Tribus, M., & McIrvine, E. C. (1971). Energy and information. *Scientific American, 224*, 179-188.

Wasserman, G. D. (1955). Some comments on methods and statements in parapsychology and other sciences. *The British Journal for the Philosophy of Science, 6*(22), 122-140.

Wittgenstein, L. (1958). Philosophical investigations (3rd ed.; G. E. M. Anscombe, Trans.). New York: Macmillan. (written 1945-1949, published posthumously)