COMPARISON OF RULE BOUND SYSTEMS THEORY TO TRADITIONAL SYSTEMS THEORY

PAUL A. BALLONOFF

Ballonoff Consulting Service, Washington, DC, USA

The purpose of the study of human culture has been called the study of human survival. It is almost self-evident that the single most important fact to a system of marriage is the survival of the population. Certainly many other human cultural systems, such as political and legal systems, have this same objective. Indeed, continued survival of any self-organizing system that must self-replace its membership by reproduction has the same property. Therefore the most important measure of information about such systems, from the point of view of the system, is whether the probability of continued existence of the population approaches 1 after the act of reproduction is carried out on the subsystem governed by the rules in question. This seemingly simple fact explains the mathematical structure of the theory of rule bound systems. It explains why and how rule bound theory differs from traditional systems theory; why information theory, but not thermodynamics, applies to rule bound cultural systems; why mathematical groups are a typical and perhaps necessary characteristic of the rule structure of rule bound systems; why the mathematics of human cultural systems require the Stirling number of the second kind but the statistics of systems studied by traditional theory require the Stirling number of the first kind; and why study of minimal systems satisfying a rule also may describe characteristics of larger systems using the same rule.

There is a long tradition of viewing human cultures as essentially literary objects. There is also another way of looking at cultures, represented by a recent work that discussed the "myth of the primitive."

Address correspondence to Paul A. Ballonoff, Ballonoff Consulting Service, Fifth Floor, 1899 L Street, NW, Washington, DC 20036, USA.

Cybernetics and Systems: An International Journal, 27:317–326, 1996

0196-9722/96 $12.00 + .00

The work shows that many cultures were not idyllic images of romantic purity but rather involved daily struggles for individual and group survival. Indeed, because many systems are truly dysfunctional, not all cultural systems survive (Edgerton, 1992). This is also true, of course, of more modern cultural systems, as is well known in recent history especially on the level of country, political, and economic systems. The theory of rule bound systems has evolved from the effort to view such human systems rigorously (such as in Ballonoff, 1994, 1987). The result is that this theory has somewhat different properties than traditional systems theory, a fact that prior to this paper was derived and commented on but unexplained. This paper explains why the two theories differ, by deriving the differences in part as the logical consequence of interpretation of the measure used in information theory.

TRADITIONAL INFORMATION THEORY

It is thus useful to start this paper by quoting from one of the more rigorous early interpretations of information theory, that by Feinstein (1958):

> Of the difficulties which confront us when we attempt to construct a quantitative theory in which the concepts "production of information" and "transmission of information" are meaningful, two stand out at once. First, we must construct a mathematical model in which we can speak of information being produced and transmitted. Second we must assign a quantitative measure to the amount of information involved. At first glance, it might appear that the solution to the second problem would follow directly from the first. ... this is actually not the case. (Feinstein, 1958, p. 2)

The traditional development of information theory is then initiated by giving motivation as follows:

> Intuitively, we would agree that we receive information whenever we are informed of an event whose occurrence was previously not certain. Furthermore, it is reasonable that, within certain limits, at least, the more likely an event is, the less information is conveyed us by the knowledge of its actual occurrence. Ignoring for the moment this last remark, we can already introduce a certain amount

of formalism into the discussion. Let x represent the occurrence of an event... [and]... Let Ix denote the amount of information conveyed to us by the knowledge of the occurrence of x. (Feinstein, 1958, p. 2)

Traditional information theory is thus derived by defining a set of mutually exclusive events and defining the information content I of each event in terms of the probability of occurrence of the event. The measure of information over the entire system is then found as a function H of the probability-weighted information of all of those events. About this fact, Feinstein noted, "This is actually a very strong condition; in fact... it practically suffices to determine the form of H... without regard for the definition of H in terms of I." [Feinstein at page 4]. He then demonstrated the fundamental theorem of information theory, which is "The maximum information content of a source having n elements is $\log n$, and is achieved only when all elements have equal probability" (Feinstein, 1958, p. 15, Theorem 1).

This concise summary is sufficient to show where things went astray in applying this theory to rigorous study of cultural systems seeking self-reproduction, for which the most relevant information is the occurrence of an event in which the system reproduces. If there is any nonzero probability of an event in which survival does not occur (that is, in which reproduction of a system capable of following the rule does not occur), then the relevant information is reduced by the existence of this probability. In Feinstein's words, we precisely cannot "ignore the last remark," which traditional theory does in fact ignore.

TRADITIONAL INFORMATION THEORY AND THERMODYNAMICS

We can now show why information theory resembles thermodynamics and then in turn why neither thermodynamics nor information theory, as interpreted traditionally, is isomorphic to rule bound theory. As just summarized, information theory finds that the maximum information is produced when all events are equiprobable because it implicitly weights information as the deviation of the system from "randomness" defined as equiprobability of all possible events. Now consider the underlying combinatorial density function of thermodynamics. Thermodynamics requires distributions based on the Stirling number of the first kind—which is to say that thermodynamics assumes all particles of a system are identical to each other (such as that all the molecules of air in a room are similar) but that all locations into which they may be put are different (the spatial distribution of air in the room is important). It then turns out that of the possible ways in which air can be distributed in a room, that which has the highest probability (largest possibility density) is the one in which air is distributed evenly throughout all of the space in the room. This is the same as saying that the probability that of the event "an air molecule is in this part of the room over here" is identical for all parts of the room to which one might point (see van Lint & Wilson, 1992, pp. 472–473, for the relationship of the Stirling number of the first kind to the counting of identical objects into nonidentical cells). This in turn results in the first theorem of information theory, that the information is maximum when all events are equiprobable. From this formal identity, based essentially on "ignoring the last remark" of Mr. Feinstein, general systems theory infers the universality of both thermodynamics and information theory and then applies the formal structure of both to problems of cultural science.

CHOICE OF DENSITY FUNCTION

This is why one of the more interesting results of rule bound theory is that it turns out to require use of a very different underlying combinatorial density function for the counting properties of populations following a rule than required for the density functions of statistics of systems studied by traditional theory (Ballonoff, 1987, Appendix II). The reasoning is that because the underlying mathematics that relates populations to cultural configurations is a surjection, statistics based on the Stirling number of the second kind are required for rule systems. This is also mathematically identical to saying that at a critical point, rule bound theory requires computing a possibility density function based on the distribution of nonidentical objects into identical "locations." [This reasoning is similar to that commonly found in standard books on combinatorics such as van Lint & Wilson (1992, p. 106) and Grimaldi (1989, p. 178) for the relationship of surjections and counting functions to the Stirling number of the second kind.]

These, however, are the opposite assumptions from those of thermodynamics: it would be as if assuming that all atoms are unique and all locations in the room are nondistinguishable. This hardly makes sense as a description of the physical system. But in cultural rule theory, the underlying mathematics are the mappings of unique individuals onto graphs showing actual and possible relationships among identical roles and also the mappings from possible graphs of social structure in one generation to the possible graphs in the next. Unique individuals fill identical roles; the graphs are possible configurations of those roles. For the system to survive with the same culture it is critical that a configuration be reproduced that is capable of (1) permitting the functions of the culture to occur given the existence of that particular graph in that generation and (2) in turn also allowing successive generations of configurations to occur, each of which permits the operations of the culture to occur in its own generation. Although all roles of a given type may be identical, not all configurations of roles permit reproduction to continue to occur according to the rules; therefore, not all possible distributions of unique individuals into a given number of identical roles will permit the system to survive.*

INFORMATION THEORY OF RULE BOUND SYSTEMS

We can now construct an information theory of rule bound systems. Define an event as the occurrence of a particular configuration of role relationships. Define the information content of a particular event as the product of whether a particular configuration permits the system to survive (1 if yes, 0 if no) times the relative probability that the rule permits the given graph to form (out of all possible graphs allowed by the rule in a population of a given size). This information is maximized over the whole set of possible events when the rules maximize the creation of configurations of role structures that can survive and reproduce (hence permit continued existence of the same rules) and minimize the probability of creation of graphs that cannot do these things. The practical meaning of the statement that cultural theory uses surjections is that an algebraic map representing graph formation under a rule, from possible configurations in one time period to those possible under a viable rule in the next, does not permit the formation of all theoretically describable possible configurations or maps among such configurations between generations. Under the most viable rules, the

*This is a verbal interpretation of the theory of minimal structures, found in Appendix I of Ballonoff (1987).

configurations that will not permit the rules to continue to reproduce will also not be formed at all or will be formed with lower probability. This means that the average information content (weighted average probability of system survival) of the configurations that are allowed to be produced is higher under what we clearly regard as the more viable rules.* Therefore also, we can construct an information theory that is consistent with rule theory—in fact, I just did it verbally.

But this version of information theory is not identical to thermodynamics. In particular, the information theory consistent with rule theory does not contain the vaunted first theorem, that information is maximized by the equiprobability of all possible events! That this is true is already apparent from the definition of information used above. In the rule bound version of a "first theorem" we need instead a limit formulation that states something quite different: as the population size approaches (from above) the size of the minimal structure that permit the rules to survive, then the information is maximized for the rules that cause a greater proportion of all possible configurations to be isomorphic to the minimal configuration. That this is true follows from the definition of a minimal structure (minimal configuration) as the smallest that permits the rule to operate in a given generation and to survive and reproduce an isomorphic structure in the shortest possible number of generations. (All known minimal structures for marriage rules reproduce isomorphic images in one generation.)

MATHEMATICAL GROUPS IN RULE BOUND THEORY

This discussion leads naturally to the role of mathematical groups in rule bound theory. The construction of the configurations describing relationships in a particular generation consists of (at least) two different mathematical objects. One of these is the mapping of individuals onto a set of identical roles; this is the function counted above by the second Stirling number. The other is a function describing relationships (configurations) among these roles in the minimal configuration; this is the structure that has been described in previous rule-theory papers and

*For example, rules that produce 1-stable configurations in their minimal sructures as defined in Ballonoff (1987, Appendix I) have very high information content by this measure—in fact, in a system that can reproduce exactly the minimal structure in configuration each generation, the information content is 1; survival is assured.

other works in anthropology as being a mathematical group—usually a symmetric or permutation group. Following from the above, the significance of these groups and therefore of the various possible mathematical representations that can be derived from the analysis of configurations by groups is that these minimal conditions assure survival of the rule, thus also of the culture.

In particular, the minimal structures are described by symmetric groups whose group order is one-half of the population size of the number of individuals required to fill the minimal configuration (Ballonoff, 1987, Appendix I). The symmetric groups also describe labelings of the relationships in the configuration (such as, in marriage theory, clan names or family names). In fact, these labelings and the rules of their transformation over time under the operation of performance of the culture can be considered as a literal message transmitted over time by the culture. In many cultures, many different cultural subsystems (such as not only family names but also religious ideologies, mythological systems, and so forth) are literally isomorphisms of the same labeling system. As the system reproduces, these messages are also continually repeated. But now consider that several different nonidentical labelings can be made of the same isomorphism class of configurations; this is true in part because several or even many different partitions can be made from the same set of labels, each of which could conform to application of the rules of a particular cyclic permutation group mapping over a sequence of isomorphic configurations. If the culture is taken as the set of acceptable labelings, this implies that more than one nonisomorphic culture can be described by the same isomorphism class of underlying configurations. Otherwise stated, several different cultures can have similar population statistics. Otherwise also stated, it is possible that there are cultures that have identical minimal structures but very different literary descriptions.

REAL SYSTEMS LARGER THAN MINIMAL REPRESENTATIONS

It is worth noting that for cultures that in fact treat the labeling as a literal "cultural message," assuring that each generation creates a minimal configuration therefore also literally maximizes the transmission of the cultural message with least error. This occurs in part because, on a system of minimal size, the symbol or labeling set is also a closed finite set, which is transmitted as messages across generations according to unique rules represented by the group operators. The message repeats as long as the population survives biologically.

Now notice that not all configurations larger than the minimal structure also assure that the rule will survive. For example, if a larger than minimal population all happened to be configured into relationships that could not reproduce under the rule, then added population size would not add to survival probability. This implies that in a larger than minimal population, for the system to be viable, the configuration in each generation must be reducible in some sense to a configuration that is viable under the rules. We know that this will be a configuration no smaller than the minimal configuration.

With this background we can now apply the mathematical theory of group representations, and in particular of irreducible representations and of group characters (traces of matrices), to get very useful results. First, except for the trivial culture containing no rule, hence using the one-element group as the representation of the minimal structure, the group describing the minimal representation of a rule has trace = 0. This must be true because if the permutation group describing a rule contains a nonzero trace, this refers to a group element that interacts only with itself, and therefore follows a rule different from the rest of the culture, and therefore follows a different rule. (In the known mathematically described cultural systems, this also violates the least necessary "incest" prohibition of self-marriage in the nuclear family.) Therefore, no matrix with a nonzero diagonal element can be the permutation matrix of a minimal representation. Therefore, the trace or characteristic of the minimal configuration of any rule is zero.

This now permits application of the representation theory for groups, such as laid out in Ledermann (1989, especially p. 24). It is shown there that if a group representation has trace zero and is the minimal representation, all larger representations can be reduced to the unique minimal one, and that the sequence of reduction from the larger to the smaller representation does not affect this result. (To be more precise, because the rule is described by a finite group of characteristic zero, every matrix representation of that group is completely reducible, known as Maschke's theorem, and then by Schur's lemma all such representations are equivalent.)

This is an extremely powerful result for cultural theory. It means that the minimal structure for a cultural rule and all of the culturally relevant labelings and messages related to that structure by isomorphisms and similar mappings are all also relevant to describing and predicting properties of any population of larger size using the same culture. This therefore also explains why the population statistics derived in Ballonoff (1980, 1982) are applicable not only to minimal systems but to systems of any size using the cultural rules, as was also stated but without proof in those papers. It also explains why analysis of cultures as the seemingly idealized "simple" configurations such as often found in myths or tales, or as described in the ethnographies of social anthropologists, in the minimal structures of rule bound theory, or equivalently in the group of operators of any rule bound system postulated by the "group axiom" of Ballonoff (1994), is actually a very powerful tool for analysis of any real culture and real population using such culture.

CONCLUSION

The foregoing analysis follows from a different definition of information at the outset than is used in traditional systems theory. The results, however, are very strong. Traditional systems theory does not explain either the origin or the persistence of human cultures, or even of living systems generally. But use of an information theory that measures information as related to survival probability leads to a complete, coherent, and highly effective predictive theory of human culture. The theory can probably also be generalized to any living system.

REFERENCES

Ballonoff, P. 1980. EMCSR '80, published 1982, Mathematical demography of social systems. In *Progress in Cybernetics and Systems Research*, Vol. X, ed. R. Trappl, 101–112. Hemisphere Publishing Corporation, Washington.

Ballonoff, P. 1982. Mathematical demography of social systems. In *Cybernetics and Systems Research*, Vol. XI, ed. R. Trappl, 555–560. North Holland Publishing Co.

Ballonoff, P. 1987. Mathematical Theory of Culture. Monograph of the Austrian Society for Cybernetic Studies, Vienna.

Ballonoff, P. 1994. Theory of rule bound human systems. *Cybern. Syst. Int. J.* 25:837–860.

Edgerton, R. 1992. *Sick Societies, Challenging the Myth of the Primitive*. New York: Free Press.

Feinstein, A. 1958. *Foundations of Information Theory*. New York: McGraw-Hill.
Grimaldi, R. P. 1989. *Discrete and Combinatorial Mathematics*. Reading, MA: Addison Wesley.
Ledermann, W. 1989. *Introduction to Group Characters*. Cambridge, UK: Cambridge University Press.
van Lint, J. H., and R. M. Wilson. 1992. *A Course in Combinatorics*. Cambridge, UK: Cambridge University Press.