Intelligent Data Analysis of Intelligent Systems

David C. Krakauer, Jessica C. Flack, Simon Dedeo,
Doyne Farmer, and Daniel Rockmore

Santa Fe Institute, Santa Fe, New Mexico, 87501, USA

Abstract. We consider the value of structured priors in the analysis of data sampled from complex adaptive systems. We propose that adaptive dynamics entails basic constraints (memory, information processing) and features (optimization and evolutionary history) that serve to significantly narrow search spaces and candidate parameter values. We suggest that the property of "adaptive self-awareness", when applicable, further constrains model selection, such that predictive *statistical models* converge on a systems own *internal representation* of regularities. Principled model building should therefore begin by identifying a hierarchy of increasingly constrained models based on the adaptive properties of the study system.

1 Introduction

The explosion of large, well curated data sets in the biological and social sciences, coupled to an exponential increase in computing power, has generated a rather novel predicament for quantitative science. This is the prospect of sophisticated analysis coupled to remedial comprehension. The traditional bias towards compact descriptions of data, ideally founded on an understanding of microscopic dynamics and interactions, has until recently, served as the Gold standard for scientific insight. Maxwell's equations, Newton's theory of color, Einsteins theory of gravitation, Mendel's "laws" of chromosomal segregation, Boltzman's contributions to statistical mechanics, and Turing's theory of computation, are all representative of this traditional construction. They all share the very desirable property of being graspable by a single mind as a coherent unity, and provide both predictive, and in some cases, engineering insights. We might call this intelligent approach to empirical phenomena - intelligible data analysis. In contrast, the proliferation of visually stunning illustrations of high dimensional data sets, that often serve as the conclusion of quantitative analyses, represent intelligently analyzed, unintelligible data analysis.

So what are we to do? The luxury of assuming simple, typically linear, relationships between elements of a small set of observables, is no longer something we can afford. Much of the adaptive world does not present itself in this form. This should not justify ceding prediction to powerful computers, except in those cases where scientific explanation is not the priority - for example, medical diagnosis or industrial design. In this contribution, we argue for intelligible data analysis realized through the appreciation of key properties of intelligent, or

perhaps more accurately, adaptive systems. We suggest that adaptive dynamics implies a set of constraints, on both function and mechanism, that can profitably be used to constrain prior beliefs. With these constraints we explore a far smaller number of dimensions than is possible under effectively unlimited processing power with a reduced sensitivity to noise. We can make this claim stronger by recognizing that for many adaptive systems, intelligible data analysis of their own states is at a premium, and so an elementary form of "adaptive self-awareness" or self-observation imposes further limitations on the complexity of a model. Stated differently, if our statistical model for behavior is accurate, we might expect it to converge on the system's own model for behavior.

In summary our principles are as follows:

1. The results of data analysis should be intelligible and not merely predictive.
2. Properties of adaptive dynamics impose cognitive constraints - structured priors - on the space of degenerate explanations.
3. Elementary "adaptive self-awareness" or self-observation imposes further constraints on these solutions.

2   From Kepler's Curves to the Machine-Read Mind

In the following two examples we illustrate on the one hand issues of comprehensibility, and on the other hand, issues relating to the value of a priori system knowledge. In the first example from the 16th century, we discuss how limited computing power can lead to more parsimonious representations of data - celestial orbits. In the second example from our own century, we discuss how classification using machine learning techniques often tells us little about how a cognitive system performs discrimination, when it ignores system mechanics.

2.1   Sensible Bounds on Computability

The contributions of Copernicus and Kepler to our understanding of celestial motion are well known. These are worth summarizing, as they epitomize a canonical mode of intelligent data analysis based on compressed descriptions of carefully curated regularities. From his observatory, Copernicus (1473-1543) made a number of important observations and reached several important insights: (1) The retrograde motion of the planets is a result of the relative motions of the earth, (2) the rotation of the earth is responsible for the apparent rotation of the stars, (3) the annual cycle of the sun is caused by the earth rotation around the sun, and (4) all celestial motion is circular and hence the apparent motion of the planets is based on epicycles moving along larger circles – deferents - neither of which need be orbiting a larger mass.

Kepler (1571-1630) worked on the Mars orbital data provided by Tycho Brahe. After consistently failing to fit Martian motion using circular orbits and ovoids, Kepler determined the path compatible with a simple ellipse. This idea was generalized to all planetary orbits with the sun occupying in each case one focus.

Subject to the limitations of measurement, the Copernican circle was not in all cases demonstrably worse than the Keplerian ellipse. Moreover, Kepler had no mechanical model from which the ellipse could be derived, and hence the criterion for its acceptance was simply parsimony.

In more modern terms, assuming the data to be elliptical, we need a rather unintelligible, infinite Fourier sum (epicycles) to provide a perfect fit. Hence the key to an efficient representation of the data was a rational choice of basis function. Kepler had intuited an appropriate function space in order to derive a compact representation of the data. If Kepler had had access to super computers, and subject to the limits of resolution of the data, there would have been little need to forego the epicycles. It is only later, with Newton and the inverse square law, that the ellipse is provided with foundational support.

The conclusion of this case study is that finite computational power, when coupled to accurate data, can be a powerful selection pressure on model development, selecting against unintelligible, high-dimensional representation. If Copernicus had a sufficiently powerful computer, Kepler and his parsimonious ellipses might have been out of a job.

2.2   Sensible Bounds on Information

Recently some progress has been made in the once questionable pursuit of mind reading. In particular, the extent to which functional magnetic imaging can be used to infer subjective states of mind. The approach is to use appropriate statistical algorithms - support vector machines - to classify brain states upon varied but systematic stimulation.

Kamitani and Tong [17] introduced the idea of "ensemble feature selectivity" - whereby very large populations of cells (V1-V4) in the visual cortex represent each of eight possible stimulus orientations using linear combinations of activity patterns. In a subsequent study, Hanson and Halchenko [18] used activity data from the full brain of ten subjects on two data sets - houses and faces - and using statistical classifiers observed disjoint networks of brain areas diagnostic for either face or house stimuli.

In both studies, the classification did not rely on unique activity at the microscopic level, but statistical averages at the population level. The value of these studies is to illustrate the nature of large-scale statistical localization involved in perception. These projects do not make use of a model or theory of neural activity that constrains the choice of statistical model – the goal is observer discrimination.

The conclusion of this case study is that powerful data analysis tools can be used to reveal consistent correlations between physical and subjective experiences, but that without some form of structured prior, these rely heavily on global information unlikely to be available to the system itself: the experimentalist can see the whole brain, but this is not something we should assume that the discriminating machinery of the brain can see. Consideration of connectivity, for example, might allow for clustering using a more confined function space.

3   Properties of Adaptive Systems

The following brief case studies represent instances of data analyses where prior knowledge of adaptive functions modify those structures and models we use in their analysis. It is our contention that knowledge of these structured priors allows us to place sensible bounds on energetic, computational and memory capacities. These restrict the search space of models according to mechanical principles relevant to the system itself, and will tend to increase the predictive power of models out of sample by eliminating sensitive dependence on high dimensional representations. We seek to avoid problems of over-fitting of data, illustrated through epicycles, or allowing access to global information, assumed in the FMRI studies. These examples span biology, social systems and cultural artifacts.

4   Case Studies in Intelligible Analysis of Adaptive Systems

4.1   Conflict in Animal Society

Conflict plays a critical role in the evolution of social systems -both positive and negative - and is typically manifested in the form of fights between individuals over the course of their lifetimes. Much is understood about control mechanisms [19], [6], [7], [22], factors driving escalation of pair-wise contests [8], [9], [10], [11], the influence of third parties on conflict outcome through coalition formation [12], [13], audience [14], [15], reputation effects [16], and redirected aggression [21]. Less is understood about the causes of conflict, and very little understood about the dynamics of multiparty conflict –conflicts that spread to involve more than two individuals and come to encompass a sizable fraction of a group.

The standard methods for analysis of conflict derive from game theory. Game theory seeks to identify normative strategies that maximize some measure of payoff in the face of uncertainty. Games are classified according to payoff structures (constant sum, non-zero sum, symmetric etc), representation (extensive or normal form), information sets (imperfect, perfect) and solution concepts (weak and strong Nash, evolutionary stable strategy etc). The work of Thomas Schelling is perhaps the best known on conflict, and includes the idea of the focal point of a coordination game - a maximum payoff solution that is independently salient to all non-communicating players.

The problem with game theory is that it provides no quantitative framework for the analysis of large bodies of data relating to strategic interactions. To remedy this deficit we sought to invent an inductive game theory [23], or set of inferential models describing individual decision rules capable of explaining features of strategic time series data. In order to make analysis intelligible and respect the principle of "system self-awareness", we considered a cognitively structured strategy space indexed by memory capacity and coordination load.

The data are as follows. A time series of fight bouts of a given duration alternating with peace bouts. The data are derived from large populations of non-human primates in captivity. Fights-bouts have a membership that is a subset

of the total population, whereas peace-bouts include the entire population. We seek to predict membership in a given fight-bout by using information present in previous fight bouts. Hence for any pair of individuals in consecutive bouts A and B, we might want to know the probability that B engages in a fight following a fight involving A. The probabilities will vary for different pairs of individuals. In order to remove time-independent effects on individual participation in fights, we compute $\Delta P(A \to B)$; the difference between the null-expected P and that measured from the data on the number of fights (N) involving B following fights involving A:

$$\Delta P(A \to B) = \frac{N(B|A) - N_{\mathrm{null}}(B|A)}{N(A)}, \tag{1}$$

where $N_{\mathrm{null}}(B|A)$ is the average from a large Monte Carlo set of null models generated by time-shuffling the series but not shuffling identities within fights. This model can be extended to include high order correlations.

$$\Delta P(AB \to C) = \frac{N(C|AB) - N_{\mathrm{null}}(C|AB)}{N(AB)}, \tag{2}$$

as can $\Delta P(A \to BC)$, the extent to which the presence of an individual at the previous step predicts the presence of a pair at the next step:

$$\Delta P(A \to BC) = \frac{N(BC|A) - N_{\mathrm{null}}(BC|A)}{N(A)}, \tag{3}$$

These probabilities can be reinterpreted as decision rules that live in the space of two-step correlation functions that vary in the demand placed on memory and the computation of coordination. This class is illustrated in Figure 1. Independent psychological testing and behavioral observation restricts the subset of candidate rules to a few tuples towards the apex of the lattice. Using these simple rules, with suitable coarse graining over individual identities (where A might be some class of indistinguishable individuals), we are able to predict - generate a forward model - for collective features of the population dynamics, including the distributions of fight bouts – consecutive fights that are propagated until termination by a given choice of probabilistic rule set. Hence antecedent knowledge of primate cognition restricts the structure of the statistical model. In so doing, we analyze within an intelligible low-dimensional space of solutions, and respect the endogenous computational properties of the sample population.

4.2   Unintelligent Agents in an Intelligent Economy

As originally observed by Adam Smith, the economy is a complex system in which agents interact with one another to produce aggregate behavior whose emergent behaviors are qualitatively different from those of the individual actors. The mainstream approach to economics over the last fifty years has focused on the strategic interaction of rational agents, typically assuming that the behavior of agents can be modeled as the Nash equilibria of an economic game. Many criticisms of this approach have been made [24]. For example, is it plausible to imagine that real
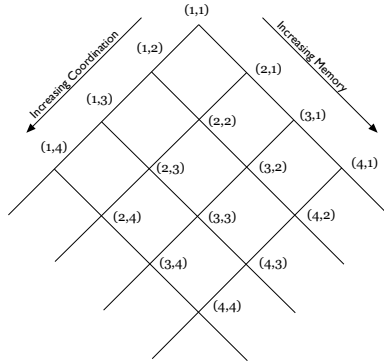
Fig. 1. Structured cognitive prior for strategy space. All strategies live in the space of 1-step Markov transition functions. Starting with the simplest model class $\mathcal{C}(1,1)$, we can add individuals to either the first or second fight, systematically building up strategies of increasing complexity based coordination, and memory constraints.

agents possess the information sources and information processing to do the difficult computations needed to find the equilibria? Perhaps even more detrimental to economics is the fact that for the modeler, since equilibrium models are difficult to formulate and solve, and are generally require analytic solutions, it is necessary to make drastic simplifications. This means neglecting important aspects, such as institutional structure, or dynamic interactions between agents under changing conditions.

At the other side of the spectrum is the subfield of economics called econometrics, which consists of making purely data driven statistical models. This approach fits functions to data, typically without strong priors about the mechanisms generating such data. Many of the best predictive models in economics, such as those that make forecasts of GDP growth or unemployment, are econometric models. This approach suffers, however, from a lack of clear interpretation. It is predictive but not intelligible.

An alternative is to abandon the quest to incorporate the strategic interactions of agents and instead focus on the institutions that structure their behavior. The modern economy has culturally evolved, or engineered, a diversity of instruments to facilitate trading. Hence economic actors and economic markets have co-evolved to solve problems of pressing economic concern, such as the institutions through which goods are bought and sold. The auctions used in the nineteenth century, such as those modeled by Walras, were cumbersome and slow. Modern markets, in contrast, typically employ the continuous double auction, which has the advantage of being fast and convenient to use. Transactions can be made instantaneously and the user need only place trading orders for a given price and volume, rather than needing to compute an entire demand schedule, as was done in the 19th century Paris Bourse.

In zero-intelligence models, agents are assumed to make their decisions more or less at random [25]. They must nonetheless respect the structure of the auction. This means that if the bid placed by a buyer is at a price equal to or greater than

the best offer by a seller, there will be an immediate transaction; if not, the bid will sit in an order book, waiting for a seller to make an offer at a lower price. Buyers and sellers can cancel their standing orders at any time. Zero intelligence models have shown that this imposes strong constraints on possible outcomes. Thus, knowing the rates at which orders flow in and out of the market, it is possible to predict properties of prices, such as the volatility (the rate of which prices vary), as well as the average spread (the typical difference between the lowest selling price offered and the highest buying price bid) [26,27].

In this case the prior is incorporated by explicitly addressing how the structure of the market mechanism for making trades shapes outcomes, as reflected in the price. Zero intelligence models can be improved by taking some of the strategic properties of the agents into account, either by empirically characterizing the behavior of real agents, or by imposing the principle of market efficiency, which states that no agent can make riskless profits [28]. It is thus possible to build up from simple models to more sophisticated models , which can make rather striking predictions about the relationship between one set of state variables and another.

4.3   Authenticating Art with Visual Priors

The art object is a product of cultural coevolution – material culture evolving upon the constraints of an evolved perceptual system. The content of figurative or landscape paintings are identifiable by virtue of key shared features with natural scenes and objects. Above and beyond matters of accurate representation, there is individual variation in style. How is recognition of subject maintained alongside a signature of the author.

The process of art authentication – the determination of the creator of a work of art – is a complicated one, accomplished primarily through connoisseurship (the opinion of an expert on the artist in question), often supplemented by various kinds of physical evidence derived from the analysis of the materials used to create the work in question. The connoisseur brings a deep familiarity with the artist, an encyclopedic knowledge of the corpus of work, and experience and information that helps him or her rule on the consistency of the work in question within the body of work. Phrased in this way, the problem of authentication takes on a statistical cast and recently, with the advent of easily obtained high fidelity digital representations of visual art, progress has been made on making sense of a quantitative and statistical approach to authentication (see e.g., [3,2]). This line of work is known as *visual stylometry.*

The methods that have been developed for visual stylometry are varied, but of particular relevance for this paper is recent work that uses the biologically-inspired method of sparse coding [1]. Sparse coding is an idea first developed by Olshausen and Field to understand visual coding [4,5]. The idea is based on the fact that "natural scenes," i.e., those scenes that our visual system encounters every day, have a certain (non-random) statistical structure. Given that, Olshausen and Field hypothesize that the visual system would have evolved to take advantage of that structure in the sense that the structure of the receptive field characteristics in

the visual cortex would optimally encode a randomly sampled natural scene. The visual system has evolved a strong prior expectation of the structure of the natural world.

Translated to the domain of visual stylometry, the sparse coding approach takes the form of basis construction for patches of a given size (say $n \times n$) so that a random patch of a (digitized) work of art by a given artist can be reconstructed by using only relatively few basis functions. As applied to the problem of authentication, we start with a corpus of digital representations of secure works by the artist in question. Having fixed the patch size at $n \times n$ we start with the standard basis of n^2 single pixel filters. A training epoch consists of choosing from the secure corpus a random image and then from this, a random collection of $n \times n$ patches. Weights for any given image are chosen by minimizing a natural cost function that trades off between information preservation and sparsity, and having done this. These weights then provide the information from which the basis is updated. Details of this can be found in [1]. Therein, the landscape drawings by Pieter Bruegel the Elder (1525–1569) were analyzed. The secure corpus consisted of eight drawings and, 2048 patches were chosen and 1000 epochs were run in order to find the sparse coding basis. In this framework, authentication is transformed into a question of is a random patch of a questioned work, on average as sparsely represented as a secure work? It turns out that with respect to a set of secure Bruegel drawings and other drawings once attributed to Bruegel, but now no longer believed (by the connoisseurs) to be authentic, this kind of classification framework works well, generally finding practically infinitesimal p-values to support the expert opinions.

Hence rather than apply ad hoc methods of spectral analysis to a field of heterogeneous pixel intensities, a result that tends to exaggerate the dimensionality of the incoming data, one employs insights form the fact that art works both mimic the statistics of natural scenes and have culturally evolved to stimulate organically evolved visual systems. This form of intelligent data analysis renders intelligible results in a form congruent with those natural to evolved mechanisms of perception.

5   Conclusions

In this short paper, we proposed the use of the adaptive property of many complex systems, to short circuit the apparent high dimensionality of data sampled from complex system data sets. We suggest that memory and computing limitations, as well as evolutionary trajectories, can be useful sources of information for structuring a priori constraints on statistical models. In a subset of complex systems, "adaptive self awareness" implies that experimental priors can be informed by the system's own priors. This is because adaptive systems need low-dimensional information about their own states in order to predict and control variables of selective consequence.

We introduced three case studies where these considerations have proved to be of value. The first involved the analysis of conflict time series using probabilistic

models in the framework of inductive game theory [23]. The cognitive limitations of the agents placed strong restrictions on the space of plausible decision rules for explaining collective dynamics. The second was the analysis of the double auction mechanism for trading. This mechanism includes significant memory and coordination properties – allowing that very simple Markovian agents are sufficient to account for macroscopic variables such as price-impact functions and volatility. Hence assumptions of infinite memory and computing power required under perfect rationality are eliminated. The third is the use of sparse coding algorithms for authenticating artworks [1]. These algorithms make use of evolved properties of visual systems that exploit redundancies in natural scenes. This significantly reduces the dimensionality of the feature vectors required to cluster paintings.

We see the future of computer assisted scientific research as cleaving to some of the early aspirations of artificial intelligence research. Rather than use Moore's law to solve problems by brute force with little comprehension of the target of analysis (chess is of course the iconic example), use knowledge of the mechanics of the system to constrain the choice of learning model. We leave the last word to artificial intelligence pioneer John McCarthy:

> *"Chess is the Drosophila of artificial intelligence. However, computer chess has developed much as genetics might have if the geneticists had concentrated their efforts starting in 1910 on breeding racing Drosophila. We would have some science, but mainly we would have very fast fruit flies."*

References

1. Hughes, J.M., Graham, D.J., Rockmore, D.N.: Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. Proceedings of the National Academy of Sciences USA 107, 1279–1283
2. Hughes, J.M., Graham, D.J., Rockmore, D.N.: Stylometrics of artwork: Uses and limitations. In: Proc. SPIE: Computer Vision and Image Analysis of Art 7531 (2010) (in press)
3. Johnson Jr., C.R., Hendriks, E., Berezhnoy, I., Brevdo, E., Hughes, S., Daubechies, I., Li, J., Postma, E., Wang, J.: Image processing for artist identification – computerized analysis of Vincent van Gogh's painting brushstrokes. IEEE Signal Processing Magazine, Special Issue on Visual Cultural Heritage 25(4), 37–48 (2008)
4. Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609 (1996)
5. Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: A strategy employed by v1. Vision Research 37(23), 3311–3325 (1997)
6. Clutton-Brock, T., Parker, G.: Punishment in animal societies. Nature 373, 209–216 (1995)
7. Frank, S.: Repression of competition and the evolution of cooperation. Evolution 57, 693–705 (2003)
8. Clutton-Brock, T., Albon, S.D., Gibson, R.M., Guinness, F.E.: The logical stag: Adaptive aspects of fighting in red deer (cervus elaphus l.). Anim. Behav. 27, 211–225 (1979)

 9. Maynard Smith, J.: The theory of games and the evolution of animal conflicts. J. Theor. Biol. 47, 209 (1974)
10. Parker, G., Rubenstein, D.I.: Role of assessment, reserve strategy, and acquisition of information in asymmetric animal conflicts. Anim. Behav. 29, 221–240 (1981)
11. Taylor, P.W., Elwood, R.W.: The mismeasure of animal contests. Anim. Behav. 65, 1195–1202 (2003)
12. Mesterton-Gibbons, M., Sherratt, T.N.: Coalition formation: a game-theoretic analysis. Behav. Ecol. 18, 277–286 (2006)
13. Noe, R., Hammerstein, P.: Biological markets. Trends Ecol. Evol. 10, 336–339 (1995)
14. Johnstone, R.A.: Eavesdropping and animal conflict. P. Natl. Acad. Sci. USA 98, 9177–9180 (2001)
15. Covas, R., McGregor, P.K., Doutrelant, C.: Cooperation and communication networks. Behav. Process. 76, 149–151 (2007)
16. Nowak, M., Sigmund, K.: Evolution of indirect reciprocity by image scoring. Nature 393, 573–577 (1998)
17. Kamitani, Y., Tong, F.: Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8(5), 679–685 (2005)
18. Hanson, S.J., Halchenko, Y.O.: Brain Reading Using Full Brain Support Vector-Machines for Object Recognition: There Is No ÒFaceÓ Identification Area. Neural Computation 20, 486–503 (2008)
19. Dreber, A., Rand, D.G., Fudenberg, D., Nowak, M.A.: Winners don't punish. Nature 452, 348–351 (2008)
20. Taylor, P.W., Elwood, R.W.: The mismeasure of animal contests. Anim. Behav. 65, 1195–1202 (2003)
21. Kazem, A.J.N., Aureli, F.: Redirection of aggression: Multiparty signaling within a network. In: McGregor, P. (ed.) Animal Communication Networks. Cambridge University Press, Cambridge (2005)
22. Flack, J.C., Girvan, M., de Waal, F.B.M., Krakauer, D.C.: Policing stabilizes construction of social niches in primates. Nature 439, 426–429 (2006)
23. Dedeo, S., Krakauer, D.C., Flack, J.C.: Inductive game theory and the dynamics of animal conflict. Plos. Computational Biol. (2010)
24. Farmer, J.D., Geanakoplos, J.: ÒThe Virtues and Vices of Equilibrium and the Future of Financial Economics. Complexity 14, 11–38 (2009)
25. Gode, D.K., Sunder, S.: Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. The Journal of Political Economy 101(1), 119–137 (1993)
26. Daniels, M.G., Farmer, J.D., Gillemot, L., Iori, G., Smith, E.: Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. Physical Review Letters Article no. 108102, 90(10) (2003)
27. Farmer, J.D., Patelli, P., Zovko, I.I.: The Predicitive Power of Zero Intelligence in Financial Markets. PNAS USA 102(11), 2254–2259 (2005)
28. Bouchaud, J.-P., Doyne Farmer, J., Lillo, F.: ÒHow Markets Slowly Digest Changes in Supply and Demand.Ó. In: Hens, T., Schenk-Hoppe, K. (eds.) Handbook of Financial Markets: Dynamics and Evolution. Elsevier/Academic Press (2008)