CRYPTO ARCHITECTURE: NOTES ON MACHINE LEARNING AND DESIGN ROBERTO BOTTAZZI

I

ROBERTO BOTTAZZI is an architect, researcher, and educator based in London. He studied in Italy and Canada before moving to the UK where he is the director of the Master in Urban Design at the Bartlett, UCL. His research analyses the impact of digital technologies on architecture and urbanism. He is the author of *Digital Architecture beyond Computers: Fragments of a Cultural History of Computational Design* (Bloomsbury, 2018) and one of the editors of *Walking Cities: London* (Camberwell Press, 2017).

In his concluding remarks on the state of digital architecture, Antoine Picon¹ advocated a return to some forms of opacity in digital design, a discipline that in his view had wholeheartedly embraced transparency as an ethos and method. In aligning digital architecture with modernism, Picon saw digital design dangerously becoming a direct, transparent expression of its own working methods which did not need critical reflection to validate its status. This task might have appeared increasingly more urgent at the time giving the simultaneous diffusion of Big Data techniques for analysing massive datasets with its promise to re-shape all aspects of society from social media to financial markets.

Two years earlier, Chris Anderson had in fact provocatively announced that statistical analysis of massive datasets would have made theory redundant.² The article elevated Big Data to a new paradigm for scientific disciplines, a transformation that humanities were also quick to follow. Anderson's argument is a well-known one by now: The power of statistical analysis engendered by evermore powerful computers and algorithms would put an end to the scientific method as modern science had introduced it in the seventeenth century. Hence, the substitution of theory with statistical correlation. On the one hand, Anderson's argument has only strengthened over time as machine learning algorithms provide an effective method to deal with both even larger datasets and forms of cognition beyond human ones. On the other, the risk of this dual transformation is to uncritically embrace transparency as a conceptual model: In urban research, for instance, the promise is that by amassing data, parsing them through algorithms, one might generate entire designs bypassing any reference to historical or disciplinary concerns. In this scenario, only a greater, more complex theoretical and historical framework—rather than its abdication—would enable us to discuss the work. Casting aside such a *Semio-clast* attitude—as Paolo Fabbri puts it³—the argument proposed here is to open up such conversation to see the relation between data and design as a constant process of writing one into the other. The data deluge animating the urban environment is re-written into more or less tangible or permanent elements of the built environment: Similar to the mathematical term re-writing, a method for substituting (potentially even in a non-deterministic fashion) the sub-terms of a formula whilst maintaining mathematical identity, data can be engrained into design through a process of continuous substitution; a process that allows for invention and playfulness providing a

1 Picon, Antoine, *Digital Culture in Architecture: An Introduction for the Design Profession*, (Basel: Birkhäuser, 2010).

2 Anderson, C, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', (2008), available at: <http://www.wired.com/2008/06/pb-theory/>, [Accessed on August 20th, 2018].

3 Fabbri, P, *Lezione di Paolo Fabbri sulla crittografia*, available at: <https://www.youtube.com/watch?v=b333yH9BgA4>. [Accessed on October 8, 2018].

diaphanous filter to positivist approaches. Spatially this is a problem of distribution here controlled by machine-learning algorithms managing the interplay between patterns and randomness.

Similar conversations have unfolded in architecture drawing from philosophical or biological paradigms. However, more recently, digital architects have been looking more closely to the actual architecture of computation, to its intrinsic properties. Amongst these topics, cryptography has received little attention despite being a core concept of computation since its inception. Cryptography in fact plays with transparency and opacity, with obscurity and intelligibility. Its role in history has not only been connected with diplomacy and secrecy, but also with the playfulness arising from the creation of systems of signs with a special, highly unstable relationship to meaning. Digital computing also involves a continuous process of encryption and decryption which moves signs in and out of the realm of human legibility, bouncing them between the inhuman language of machines (based on binary code for digital computers) and comprehensible texts, images, sounds, etc. This has hardly been just a technical exercise as, since its first formal definition, it has been entangled in cultural and political conversations. Generally speaking, cryptographic operations divide into transposition and substitution. When designing with a deluge of data, the only practical approach is to move to a higher, more systemic level to manipulate abstract information. Designing with data becomes a problem of (spatial) distribution and substitution (transposition is a far less transferable technique); technical operations can be charged with intellectual intent moving the parallel between cryptography and design beyond the mere metaphor. This aligns with architecture on two different temporal frames. Architectural discourse has long speculated on the impossibility to regulate or even fix the relation between signs, meaning and interpretation giving rise to computational design methods also based on strong syntactic variations and weak semantic qualities. On a broader timeframe, cryptography alludes to logical techniques to venture into unknown numerical or (for architecture) spatial domains that might lie beyond human cognition. In architectural research, this echoes with the possibility to appreciate massive scales such as that of the planet, or confront temporalities both below and above our perceptive threshold (from climate change to high-frequency trading). Cryptography has a long history of dealing with these 'inhuman' domains characterized by a complex relation between patterns and randomness, and the 'alien' dimension of massive datasets.

Cryptography also involves cryptanalysis, the process of decoding signs. Since its emergence in the sixth century in Persia, cryptanalysis exhibits complex and highly relevant characteristics for this discussion: for instance, the use of statistical distribution to seek out correlations and substituting characters. The use of 'automatic' methods to relate signs to one another is not only still in use, but also holds

the powerful idea that it is possible to manage signs without any prior knowledge of their meaning, as “messages without a code” paraphrasing Roland Barthes.⁴ This is all the more relevant when we compare these proto-cryptographic techniques to Ramon Llull’s system as presented in his *Ars* at the beginning of the fourteenth century. A device made of a series of concentric wheels generated combinations of letters that would then be decoded through a series of charts (proto-ciphers). Though the work of the Spanish philosopher derived from studies in mnemotechnics, he also confronted basic problems of cryptography such as the relation between syntax and semantics. However, the cultural and religious motivations inspiring his work prevented Llull from utilizing his wheels in what today we would call generatively: triplets and quadruples of letters randomly produced by spinning the wheels had to be discarded because, once decoded, they would give rise to sentences contradicting Llull’s own metaphysics. In other words, the device produced an open system through recursion; such a complex syntax with its unprecedented amount of complexity had to be severely curtailed by a strict semantic system which only admitted a finite and predetermined number of solutions. The reward of such contradictory choice was the preservation of Llull’s message. The relation between data—as invariant elements—and rule or, in modern parlance, algorithm—as a method to apprehend them—is crucial here as Llull’s example already clearly shows how the two connect: The rule or set of rules we use to investigate affects what is returned. The renaissance would introduce important developments on the relation between cryptography and mnemonics. Whilst the latter only accepts a close relationship between signs and meaning, the former is actually interested in expanding the relation between the two—in fact, it wants to make it as broad and different as possible! By removing meaning from the equation, cryptographers could fully exploit the possibilities engendered by combinatorial logic, allowing for a much greater number of combinations, which eventually overcame the weaknesses of statistical frequency. The first polyalphabetic ciphers were introduced by Leon Battista Alberti and would be fully refined by Blaise de Vigenère in 1586.

The decoupling of meaning and syntax (now an algebraic problem) would give rise to modern cryptography and with it to a more complex notion of ciphers. Ciphers not only refer to the instrument invented by Alberti to encrypt and decrypt information, but also to their etymology indicating the key to resolving a mystery, injecting semantics into what only possessed syntactic integrity. The cipher tunes the two domains, it is not given, it must be invented, designed, it requires a great deal of artistry whilst being solidly resting on strict syntactic rules; it

4 Roland Barthes, ‘Rhétorique de l’image’, *Communications*, 4.1 (1964), 40–51 42).

captures elements of randomness, of an unknowable dimension through rational operations of combinatorial logic. These are also the basis of coding—albeit here still in its primitive purely numerical form. For this reason, the short passage at the end of Pico della Mirandola's *Heptastichus* (1489) is of great relevance. In describing the origin of the world, Pico combines both individual letters and syllables to give rise to an open play of signifiers in search for meanings that could not be known in advance. Randomization (here understood as an 'extreme' form of combinatorial logic that can proliferate unbound) was an essential method to dislodge existing notions, to rationally venture into the unknown, the apparently irrational. As Umberto Eco⁵ beautifully described it, this marks a clear difference with Lull's exercises whose reward is to free man from the laws of the cosmos, to set it as a mobile subject negotiating his position in the world, and to affirm a humanist project not as mere celebration of mankind but rather a more complex exploration in which man could grasp and potentially alter the very laws governing its existence. Combinatorial logic opened up to embrace and, in fact, seek out new worlds with an emphasis on learning and the unknown.

In 1913 Andrei Markov presented his analysis of Pushkin's stories. It was not a traditional piece of literary criticism, but rather a study purely based on the statistical distribution of letters neatly divided into vowels and consonants. In this work, randomness and cryptoanalysis conflated time opening up a method that is still largely used in social media and translation software. Markov inverted Lull's idea of treating letters as numbers and delivered statistical and algorithmic methods to compute and translate natural languages without the aid of prior knowledge of the text or even the language used in it. The emphasis moved from discrete elements (the individual letters making up the Pushkin's text) to the pure relations between the individual signs. Relations were treated purely statistically: patterns and randomness emerged in their own terms and were mined without the need to reduce the complexity of the text. For our argument this example shows the sheer power of re-writing and the artistry required to perform such operations. Markov overlooked semantics which consequently turned the original text into a cryptic message to unravel through the mathematization of language. By proceeding through pure counting, the whole of the text could be tackled (though, for practical reasons, Markov only parsed groups of 100 letters at the time) and through substitution a different image of its distribution emerges. Through a technical expedient, a different aspect of Pushkin's work can be charged with conceptual intentions.

5 Umberto Eco, *From the Tree to the Labyrinth: Historical Studies on the Sign and Interpretation* trans. by Anthony Oldcorn (Cambridge, Massachusetts: Harvard University Press, 2014), 414.

The power of statistics also extended to design disciplines as in the 1920s when Cornelis van Eesteren and Georges Pineau joined forces to produce a competition entry for the new traffic plan of Paris in which models for demographic analysis and statistical projection were used for both physical and social data. It was, however, the introduction of analytical tools in linguistics studies that was to have a deeper effect on architectural research. The work developed in Ulm, Germany, from the 1950s by Tomas Maldonado, Abraham Moles, and Max Bense employed analytical and 'scientific' methods to study the relation between signs and meaning in the physical environment. On the other side of the Atlantic, similar pre-occupations were directing artists and, later on, architects towards more systematic studies on (computational) linguistics. Peter Eisenman deeply contributed to this conversation with his definition of conceptual architecture which builds on Chomsky's work on language as a computational, recursive system to take advantage of the instability between signs and interpretation as a means to charge architecture with greater conceptual depth.⁶ The results of these experiments were captured in the famous series of ten houses produced at the time. The riddle-like nature of the spatial organization of the houses was attained through an intricate process of formal manipulations which only constitutes the most superficial correlations to examine here. More interesting was the creative, almost playful nature of the permutations explored through the design process based on a strict logical syntax to which corresponded an open, unstable semantics. This achieved a complex, alienating quality which Eisenman considered essential to the emergence of the conceptual spatial understanding. This element resonates with the exploratory, experimental nature of signs substitution also informing current design work on machine learning in which syntactic integrity is prioritized over semantic clarity. Eisenman's game was nevertheless hermetic, constrained by form as the sole set to play with.

Such historical considerations take a different meaning once we are confronted with the emergence of fully automated, data-driven architectures like contemporary factories and distribution centres. In buildings such as Amazon Fulfilment Centres, we literally experience spaces continuously encrypting and decrypting data, which bounce back and forth between machine language and intelligible signs. Here historical discussions on the relation between data, sign, and space take a new turn. Mnemonic techniques inherited from the classical tradition to provide a system to order space give way to algorithms no longer seeking to reduce complexity to attain an intelligible order; the dream of total transparency promised by computational treatment of data is radically complicated. In distribution centres, goods can be

6 Peter D. Eisenman, 'Notes on Conceptual Architecture: Towards a Definition', *Design Quarterly*, 1970, 1.

placed anywhere as they are trackable through radio-frequency devices (RFID), movement of workers is not based on visual clues; rather paths are calculated and communicated to robots (or humans) that will retrieve a particular good. It is a space of algorithms in which the physical experience, as ephemeral as it may be, results from the complex ecology of machines and codes circulating through it: a problem of distribution. Again, the scale of the information flowing in them can be planetary and highly granular as it conveys individual purchasing orders or desires. The problem of how to relate the abstract and the physical reoccurs and so does the question of what design might mean in such conditions.

The parallel with cryptographic issues provides an interesting frame of reference. Whilst the exploration on indexing of the 1970s still allows one to read the presence of the author, here space no longer carries clear indexical signs of the transformations it underwent. The complex process linking data to space makes the relation between the two 'nonproportional and electronic'.⁷ Once this direct, one-to-one relation is lost, it is rather the notion of the cipher to link abstract to actual space in a way more rewarding than established categories of spatial analysis. Design here increasingly aligns with processes of encryption and decryption and re-writing in a rather direct fashion.

FIG 1
Chuanren Lin, Lei Wang, Qiuyang Zhang, Xi Meng. *Fluff*. Studies for a tower based on wind data. Master in Urban Design, The Bartlett, UCL, 2016. Tutors: Roberto Bottazzi, Kostas Grigoriadis.

7 N. Katherine Hayles, 'Virtual Bodies and Flickering Signifiers', *October*, 66 (1993), 69–91.

Such operations also exceed the models proposed by the coupling of biology and design or parametrics. Parametrics presupposes the identification of an ideal—no matter whether present or not—to seed out a whole series of instances all related by continuous variation. Re-writing and (more apt to spatial design) nesting allow for spatial and scalar inconsistencies; spatial articulation can take discontinuities whilst remaining logically coherent. It is no longer relying on ideals situated in a transcendental domain, but seeks pattern and randomness within its own bounds. The divergence from the biological model needs to be qualified further. It is not the emphasis on adaptation to be superseded here (on the contrary, as we shall see, this is still an important element), rather the particular way in which biological models and computational ones merged. This has mostly presupposed to move from the particular to the general: Cellular Automata describe simple conditions and rules which, once animated, can grow into complex formations. Datascape are however given ‘all at once’, not as parts and they can only be treated at the very scale they exist as, at any given point, a database is always complete and self-referential (if well constructed). Cryptographic techniques offer a more compelling framework as they allow to re-write the whole by distributing ciphers through a given domain. This also provides a departure from some of the paradigms that have guided the discourse of digital architecture in recent decades as this approach can no longer be described as top-down or bottom-up, but rather as both. That is, scalar hierarchies are no longer able to account for the jumps, connections, correlations of relational approaches delivered through learning algorithms which rest on vast, highly granular datasets. These relations can be inferred or simply tested once the algorithm itself attempts some form of categorization (unsupervised learning). No longer bound by linear scalar approaches, these processes of encryption and decryption engender greater flexibility by moving indifferently between the granular and the planetary. The type of patterns returned by the algorithmic mining of large datasets (for instance, the type we can visualize with T-SNE algorithms) are no longer reducible to Euclidean or even topological spatial models of representations as they exhibit a unique level of granularity: Clustering of similar data can be punctual or extensive; a distribution of patterns not only no longer reducible to linear classifications, which also becomes the expression of the logic of digital computation. Though most databases used in urban design are relational, this is still a fundamentally static technology. In order to work within the constraints of digital computation, at any given moment in time a database must unambiguously determine the number of items in it and their individual value. If databases are dynamic, this can only be assumed to happen in a discrete fashion. How to relate the temporality of databases to design still remains an open question requiring more experimentation. However, it is already clear that large datasets allow to operate on spatial qualities that exceed that of pure formal manipulations.

FIG 2
Anna Kampani, Apostolopoulos Apostolos, Caitling Brock. *Perceptive Datascares*. Diagrams showing formal transformations based Machine Learning algorithms evaluating visual permeability. Master in Urban Design, The Bartlett, UCL, 2018. Tutors: Roberto Bottazzi, Tasos Varoudis.

What is at stake in this work is a different form of spatial and informational distribution that could straddle opacity and transparency, jumping beyond formal coherence whilst having a robust logical basis. As complex ideas and technics increasingly penetrate design, so does the need to widen the range of fields and timeframes to avoid quick, ‘transparent’ alignments of data and design.

FIG 3
Chuanren Lin, Lei Wang, Qiuyang Zhang, Xi Meng. *Fluff*. Studies for a tower based on wind data. Master in Urban Design, The Bartlett, UCL, 2016.