ALGORITHMIC THINKING AND THE STRATEGIC USE OF ABSTRACTION

DAVE COCHRAN

Cognitive Systems Group, School of Computer Science,
University of St. Andrews, St. Andrews, Fife KY16 9SX, Scotland

E-MAIL: DAVEC@CS.ST-ANDREWS.AC.UK

KEYWORDS

algorithms, abstraction, philosophy of cognitive science, philosophy of biology, fictionalism, adaptationism, spandrels, computation, epistemology

ABSTRACT

This paper demonstrates how philosophical thought can inform science, with reference to the interplay of causal and algorithmic thinking in Cognitive Science, Evolutionary Biology and other scientific enquiries. It is argued (touching briefly on the debate between Fictionalism and Realism regarding mathematics (Field 1989; Lewis 1991)) that algorithmic concepts (eg. substrate neutrality) are mathematical fictions for describing processes in the same way that geometric concepts (eg. lengthless points, arealess lines) are for describing spatial relations. As such the act of parsing a causal process into "the algorithm" and its "substrate" is underdetermined by the underlying reality. It is demonstrated, using thought examples, that even the computational/algorithmic characterization of a physical digital computer is underspecified by the underlying causal/physical processes. In this *way it can be seen that the Dennett–Gould debate on adaptationism (Gould and Lewontin 1979; Dennett 1995; Gould 1997a & b) was not based on any real disagreement about the reality of evolution, but only the parsing of that reality into algorithm and substrate. It is shown (drawing further examples from cognitive science, phonology and physics) that the decision between differing but compatible algorithmic descriptions of natural processes is not, however, arbitrary, but rather a matter of epistemic and communicative strategy.*

INTRODUCTION

ALGORITHMIC REASONING is a method for abstracting over physical processes, and is, I will argue, ubiquitous in the special sciences; whether one is attempting to characterise language production (Levelt 1989, Cochran 2006) the learning of the speech code through audition and vocal babbling (Markey 1995, Oudeyer 2006), the firing processes of individual neurons (Koch 1999), the transcription of proteins from DNA (Crick 1958), Predator-prey cycles (Lotka 1925, Volterra 1926), or evolution itself (Dennett 1996), algorithmic descriptions, which I will argue may be shortly defined as formal abstract descriptions of processes, are invaluable as a way of drawing generalisations out of the empirical phenomenon. But how may we explicitly characterise the role played by algorithms in scientific thought? What exactly is the relationship between an algorithm and the material "substrate" that is said to instantiate it in the particular case? How may we distinguish between the two? In the course of this paper, we will see that these questions are inextricable from each other.

ALGORITHMS IN MATHEMATICS AND COMPUTER SCIENCE, VERSUS THE SPECIAL SCIENCES

In mathematics and computer science, this is often cashed out in rather more restrictive ways; as a sequence of unambiguous, explicit instructions (which may branch or loop, provided the branching/looping behaviour is also governed by unambiguous explicit instructions) for formal symbolic operations, used to solve mathematical/computational problems. There is no generally agreed precise definition of algorithm. Kleene (XXXX) characterised algorithms as finitely terminable problem-solving procedures implementable by a Turing machine or equivalent. More broadly, Markov (1954, p.1) offers the following;

> In mathematics, “algorithm” is commonly understood to be an exact prescription, defining a computational process, leading from various initial data to the desired result ... The following three features are characteristic of algorithms and determine their role in mathematics:
>
> a) the precision of the prescription, leaving no place to arbitrariness, and its universal comprehensibility -- the definiteness of the algorithm;
> b) the possibility of starting out with initial data, which may vary within given limits -- the generality of the algorithm;
> c) the orientation of the algorithm toward obtaining some desired result, which is indeed obtained in the end with proper initial data -- the conclusiveness of the algorithm.

Compare this to the line-up of salient characteristics of algorithms used by Dennett (1995, pp.50-1) to justify the characterisation of evolution as an algorithmic process;

> a) *Substrate neutrality*: … The power of the procedure is due to its logical structure, not just the causal powers of the materials used in the instantiation, just so long as those causal powers permit the prescribed steps to be followed exactly.
>
> b) *underlying mindlessness*: … each constituent step, as well as the transition between steps, is … Simple enough for a dutiful idiot … or … a straightforward mechanical device to perform.
>
> c) *guaranteed results*: Whatever it is an algorithm does, it always does it, if it is executed without misstep. An algorithm is a foolproof recipe.

It is not clear that these two definitions are equivalent or even co-extensive in all cases; in particular, the emphasis on problem-solving in computer science and mathematics definitions does not seem to apply in all cases where we might want to use algorithmic descriptions for natural phenomena; in the case of evolution, it seems at best to be a thinly-stretched metaphor, since it formulating what the *problem* is has proven frustratingly illusive. Moreover, Kleene's characterisation seems to be tied to the notion of symbol-manipulation, which makes little sense in evolutionary theory, and in cognitive science makes sense only in the context of a classical cognitivist approach.

Dennett refines this to an emphasis on "logical structure", as opposed to "causal powers", but does not specify in detail what constitutes "logical structure", except insofar as "logical" may be equated to "abstract". In fact, we can specify a little further; in order that the criterion of unambiguousness be respected, the set of entities ("variables") the interactions between which the algorithm purports to concern should comprise a rigorously defined *ontology*, which for our present purposes may be loosely conceived of as the filtering out of a manageable subset of patterns of interest from a larger “Reality”, which is itself simply to complex to be humanly understood. However, this looseness of definition might be just the thing needed for scientific practice, since it does not fix too much by definition at the beginning of one’s investigation. What this leaves us with is the notion of algorithms as unambiguous, idealised, abstract descriptions of processes; sufficiently abstract as to be multiply instantiable, possibly in diverse media. In the biology, linguistics and psychology, theories *always* concern processes, validity of theories require that they be correct of enormously diverse sets of real-world instantiations; for instance, a theory of child language acquisition should describe the emergence of language in all normal human children. Of course, theories may fail to be sufficiently unambiguous or precise to be equivalent to algorithms, but this also means that such theories struggle to yield genuine predictions beyond the vague intuitions of the theorist. Indeed, one of the great advantages of computational modelling in science is that it forces scientists to reach that high standards of precision and clarity. By this definition then, we can take it that algorithmic thought either *is* ubiquitous to the special sciences, or *ought* to be.

ALGORITHMIC FICTIONALISM

Before proceeding, I should note that in a certain sense, I regard algorithms as mathematical *fictions*. It is not relevant to the present consideration whether it is proper to regard mathematical statements as potential truth-bearers (Field 1989; Lewis 1991), but it should be noted that several mathematical concepts correspond to nothing in the physical universe; one can never count to the imaginary number *i*, one never finds a dimensionless point, a line with zero area, or a plane with zero volume, and analogously, one never finds an algorithmic process that can run substrate-neutrally for an arbitrarily long time. The Second Law of Thermodynamics requires that no algorithmic process in nature can continue indefinitely, without its causal processes eventually being interrupted or perturbed by the causal processes of its substrate. It is simply that if, say, the behaviour of an algorithm running on a computer is changed by a bug in the implementation, or a block of memory being destabilised by the machine overheating, one says "*that's* not the algorithm." An algorithm is an axiomatically defined mathematical entity which, by *fiat*, allows us to parse the totality of a natural (or computational) phenomenon into "algorithm" and "substrate".

These are, at best, abstractions to which physical systems may approximate.

Abstraction and Underdetermination

An abstraction and the reality abstracted from are always in a relationship of bidirectional underdetermination. That one algorithm maps to many possible instantiations is a commonplace, so I will present examples here to show that also, one instantiation maps to many algorithms.

Underdetermination in artefactual systems

Case 1: Take the example of a simple digital computer, processing zeroes and ones using simple logic-gate operations, doing a passable impression of a Turing machine, running a simple algorithm, let us say for the addition of integers from –10 to 10. Numbers are entered using two 21-key keypads (one key for each possible value), and the output is displayed by illuminating LEDs, from the first integer to the total, on a number line from –20 to 20. Let us say that each 0 and 1 in the computer's binary code corresponds to a range of possible microvoltages in the switches in the computer's circuits, as in fig. 1.

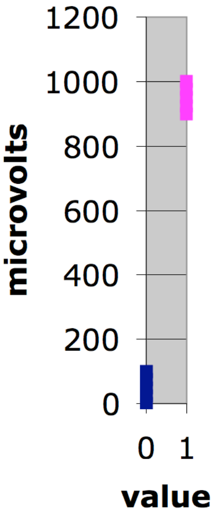

FIGURE 1. THE MICROVOLT EQUIVALENTS OF BINARY CODE IN A HYPOTHETICAL DIGITAL COMPUTER

Let us suppose that there is variation in the microvoltages for any zero or one. Let us also suppose that we can classify bits, besides being zeroes (0-100) or ones (900-1000), as "outside" (0-50 or 950-1000) or "inside" (50-100 or 900-950). This distinction makes no difference to the algorithm running on the machine; zeroes and ones are part of the algorithm, insides and outsides are mere properties of the substrate, in relation to which the algorithm is neutral.

Now let us suppose that the behaviour of insides and outsides on the machine are in fact quite regular. An inside bit in memory will remain an inside bit when copied to the processor, or to another position in the memory. An inside (i) bit inputted to a NOT gate will output an outside bit (o), and vice versa; two i's inputted to an AND gate will output an i, whereas an o and an i, or two o's, will output an o, and so on (see table 1).

gate	input	output	gate	input	output	gate	input	output
AND	i,i	i	IF	i,i	o	NOR	i,i	o
	i,o	o		i,o	i		i,o	i
	o,i	o		o,i	o		o,i	i
	o,o	o		o,o	o		o,o	i
OR	i,i	i	NOT	i	o	IFF	i,i	i
	i,o	i		o	i		i,o	o
	o,i	i					o,i	o
	o,o	o					o,o	i
NAND	i,i	o	XOR	i,i	o			
	i,o	i		i,o	i			
	o,i	i		o,i	i			
	o,o	i		o,o	o			

TABLE 1. PROCESSING OF o/i VALUES LOGIC GATES

Note that these are physically the same gates that process 0's and 1's, and the names here, AND, OR, NAND, etc, indicate the behaviour of the gate in relation to 0/1 values; thus the computer's "processing" of i's and o's is completely parallel to its processing of 1's and 0's

However, i and o values are generated randomly, as a result of thermal noise, by the input device (the keypads) and are ignored by the output device (the number line). Still, the computer itself can be seen as only implementing the “adding numbers from –10 to 10” algorithm, and i and o values are still to be regarded as algorithmically irrelevant substrate features.

CASE 2: Now, suppose the input device is replaced with one that was able to manipulate i and o values using a second pair of keypads, and the output device with one that displays the o/i output on the *y*-axis of a graph consisting of a 41x41 LED matrix, and the 0/1 output as the *x*-axis. When the computer delivers its output, it shows a vector from Cartesian co-ordinates corresponding to the first 0/1 and o/i inputs to those corresponding to the totals. Now, we would wish to say that the computer is operating with a 4-ary machine code rather than binary, and that it implements a vector-addition algorithm rather than an integer-addition algorithm. But note that this is the case even if the physical processes going on the inside the casing of the computer are *identical* to the previous case. The difference in the input and output devices, although they are *not* part of the substrate of the algorithm, has prompted us to *parse the causal process* inside the computer casing *into algorithm and substrate* differently.

CASE 3: Now, suppose the human user of the computer were to use this new set-up, but were to completely ignore the *y*-axis on the output, and enter random digits with her eyes closed on the new keypads? Which algorithm is implemented now?

CASE 4: Or what if the user were *unable* to see the new keypads (but knew to reach out to the approximate area where they are located and strike a random key, in order to get an output) and *y*-axis, due to lesions in the visual cortex? Perhaps in this case and the preceding we would say that it is the integer addition algorithm.

CASE 5: Then what about a case identical to CASE 3, but the user *remembers* the o/i inputs and outputs, and maybe uses some of them on a later occasion?

CASE 6: What about a case identical to CASE 4, but with a non-lesioned observer? Does it make a difference if the observer notes and uses the i/o inputs and outputs for his own purposes? Or for no purpose? What if he tells the user the information she has missed?

CASE 7: Or again, suppose *two* users were to operate the computer, user 1 using the original keypads and reading off the *x* axis, user 2 the new keypads and the *y* axis. Here, perhaps we should say that the computer instantiates *two* instances of the integer addition algorithm, with the algorithmic features of each a proper subset or the substrate features of the other.

In all these cases, our intuition as to the "correct" parse of algorithm from substrate has been governed by our awareness of the intentional states of human users, and the status of the physical implementer as a culturally meaningful artefact. This may strike us as slightly odd, since algorithms themselves are supposed to be mindless and mechanical. Should we then reject the application of algorithmic concepts to natural systems, where no intentionality is present? This is a move I believe should be resisted, if only because of the well-established utility of algorithmic thinking in scientific practice. We might try to fix the parsing of a natural process into algorithm and substrate in terms of its input to or output from another algorithmic process, but again the same ambiguities may arise.

Consider another kind of example. Suppose a computer is being used to run an algorithm for determining the primeness of members of a list of hundred-digit natural numbers. However, the computer contains a faulty memory chip, in which random bit-flips are caused by thermal noise against which it would be robust if it were functioning properly, and as a result, the computer gives incorrect answers. Note that here the thermal noise, which should have been an algorithmically irrelevant substrate-feature, has altered the output. However, we do not in this case say that substrate-neutrality is disproven, we simply say that the computer has malfunctioned and so failed to implement the algorithm. But now suppose, instead of ditching it and getting a new one, the owner of the computer uses it to run a genetic algorithm, designed such that the "genes" of the population are stored in the faulty chip, so that the random bit-flips provide the "mutation" part of the algorithm. Here we would say that the bit-flips are part of the algorithm, Now, suppose the owner did *not* know about the faulty chip, and implemented the genetic algorithm in the *exact* same way, except for an additional subroutine causing further bit-flips among the population's "genes". Here, the GA would still *work*, but the programmer's figures for the mutation rates on different runs of the programme would be incorrect. In this case, do we still count the bit-flips induced by thermal noise as part of the algorithm?

What these examples show is that, even in the case of the operation of digital computers, the algorithmic interpretation of physical, causal processes is not unambiguously determined by their detailed physical/causal description. Now, let us look at the consequences of such underdetermination in cognitive science and biology.

UNDERDETERMINATION IN NATURAL SYSTEMS

Now let us consider two examples of ambiguity in the parsing of *natural* phenomena into algorithm and substrate; the Dennett/Gould debate regarding the relationship between adaptation and other processes in evolution, and the case of Extended Mind.

THE STRANGE CASE OF DENNETT V. GOULD. The famously vituperative debate between Dennett (1995) and Gould (Gould and Lewontin 1979, Gould 1997) concerned the relationship between adaptation and non-adaptation in evolution, and in particular Gould and Lewontin's (1979) hypothesis that non-adaptations, such as historical accident, the recycling of obsolete systems towards new functions, and especially "spandrels", structural by-products of adaptive changes, capable of being subsequently co-opted to adaptive functions (Gould and Lewontin 1979) are just as "evolutionary" as adaptations proper. Dennett, against this, claims that "either spandrels are not ubiquitous after all, or they are the *normal basis* for adaptations, and hence no abridgement at all of pervasive adaptation" (Dennett 1995, p.268). The difference between the adaptationist and non-adaptationist perspectives is nicely revealed in Tooby and Cosmides's (1997) indignant rebuttal of arguments presented in Gould (1997a & b) to the effect that adaptationist research "neglected" non-adaptive explanations:

> In our 1982 work proposing that parasite pressure drove the evolution of sex and the maintenance of genetic polymorphism, we explicitly used neutralist theories of evolution to evaluate contrasting predictions about the distribution of alleles driven by

> chance vs. frequency-dependent parasite pressure. In this way we were and are no different than any other adaptationist biologist, who use molecular clocks and other applications of neutralism routinely in research, as well as, for example, random walk and byproduct null hypotheses.

But for Gould, a byproduct hypothesis is *not* a null hypothesis. What this seems to indicate is that the disagreement boils down to a difference in the way the parties involved parse the phenomenon into algorithm and substrate; Those in the adaptationist camp choose to interpret only adaptation by replication with mutation and selection as "algorithm"; the role of other, non-adaptive factors in the history of life is not denied, but they are regarded as substrate features and so are the null hypothesis in accounting for developments in the history of species. Gould regards non-adaptations as of sufficient interest to warrant their consideration as components of the same complex algorithm, evolution, which also incorporates adaptation by replication with mutation and selection.

Evolutionary biology, of course must proceed at a level of extreme abstraction from the detailed physico-causal facts; A wholly non-abstract account of the events the discipline generalises over would require detailed accounts of the life-histories and development of millions, billions or trillions of individual organisms, detailing the role played by each gene in each of them, their interactions with conspecifics and the larger environment, culminating in a summary of their individual reproductive success, and their net effect on the reproductive success of others in the population. In this light, and given that in either account the algorithm remains a mathematical fiction, in the sense indicated above, it seems very difficult to see what the correctness of one account and not the other would mean in terms of the totality of life-histories from which evolutionary biology abstracts away, and so the disagreement might more helpfully be understood as one about what constitutes the optimal scientific *strategy*, rather than the correct scientific *finding*.

In order to understand the sorts of strategic considerations that might influence scientists' selection of algorithmic abstractions, we will now turn to the subject of the theory of the Extended Mind.

ABSTRACTION AND STRATEGY: THE CASE OF THE EXTENDED MIND. The Extended Mind hypothesis is the proposal that the boundaries of the mind are not delimited by the boundaries of the brain, but rather, many computational processes distributed between the brain and scaffolding the larger environment are better seen as properly mental, rather than as interfaces between the mental and the non-mental. Clark and Chalmers' (1998) classic paper on the subject gives paired examples of cases where the same algorithmic process is implemented by a) the brain only, and b) brain and artefactual scaffolding; for instance;

1. In a shape-fitting task on a computer screen, rotating the shapes "in one's head" to test their fit, versus doing so with a rotate tool on the computer.
2. Searching for words in a set of scrabble tiles by shuffling them "in one's head" and physically
3. Use of one's memory to store practical everyday information, and use of a notebook by an Alzheimer's patient.

The upshot of Clark and Chalmers' argument is that the differences between the "all in the head" and "distributed" versions of each examples are "substrate" features, irrelevant to their algorithmic characterisation.

However, for the purposes of some algorithmicisations of the same processes, the Extended Mind approach may prove inappropriate, for instance if one is trying to understand the relationship between (abstract, mathematised representations of) the computational properties of single neurons (Koch 1999) and higher-level cognitive processes. Since the "extracranial" processes in 1, 2 and 3 cannot be reconstructed from such units, they could not figure in such a charaterisation.

Moreover, in examples 1 and 2 there is another way of thinking about the higher-level processes; one could take the view that the extracranial processes in these examples are separate environmental informational processes, and that one of the capacities of the algorithm/complex of algorithms implemented by the brain is the ability to *simulate* such environmental processes, in a more or less heuristic fashion. If one's investigation is of *how* the mind constructs such simulations (*qua* heuristics), or under what cognitive constraints such simulations operate (for instance, the extent to which mental Scrabble-tile shuffling is constrained by learnt phonotactic or orthographic transition probabilities), it would be appropriate to algorithmicise these processes in a way that abstracts away rather less of the detail of the implementation, to the point where the intracranial and extracranial processes appear quite to be different kinds of things.

On the other hand, attempting to understand the evolution of cognitive systems *without* the benefit of the Extended Mind perspective would be greatly impoverished, since an animal's ability to enter into cognitive couplings with features of its environment greatly alters its fitness landscape. Similarly, the epigenesis of cognitive systems is profoundly affected by the information-processing couplings they enter into with body and environment; for instance, in human infants, babbling is the process whereby infants learn the mapping between motor gestures of the articulatory tract and speech sounds (Vihman 1996, Oudeyer 2006; for analogies between babbling and "subsong" in infant songbirds see Goldstein, King and West 2003). Oudeyer (2006, pp.75-122) presents a set of models in which brain, body and environment form a closely coupled loop, where random motor commands are transformed into patterns of sound by the vocal tract (body), and are transmitted through the air (environment) to the ear and translated back into neural encodings by the electromechanical properties of the cochlea, and a mapping between the neural representations of the sounds and motor gestures is learnt. By contrasting versions of the model equipped with abstract or more realistic realistic models of the articulatory tract, Oudeyer is able to show how the biomechanical properties of the vocal tract significantly changes the structure of the phonemic inventories learnt. Here, use of an Extended Mind model (although Oudeyer does not use the term) allows for insights that could not otherwise have been reached.

The point here is that while there may, for any physical process, be multiple possible ways of abstracting algorithmic processes out of it without misrepresentation of the underlying facts, but different algorithmicisations of the same underlying phenomenon may differ greatly in the generalisations they enable, and as such some will prove superior in the context of some types of investigation, and others in others.

CONCLUSIONS

Of course, this should not be read as implying some sort of relativism; some algorithmic attempts to characterise natural processes may be poorer in generalisations than others in all contexts of investigation, or may turn out to be flatly wrong, such that the natural process is in fact not an instantiation of the algorithm at all. However, the bidirectional underdetermination between the algorithmic and the physico-causal that I have shown leaves some room for fault-free disagreement. This leaves us with the question, where there is a fault-free contradiction between two or more algorithmicisations of the same phenomenon, is it possible, or desirable to unify them? There is not time in the present treatment for a detailed investigation of this matter, but I will conclude with a moral on this point, which is that this is a question that is best answered on a case-by-case basis, on grounds of scientific expediency, rather than on the basis of a philosophical presecription; if the resolution of a fault-free contradiction results in a theory that is harder to comprehend without enabling new generalisations, it is probably better to leave well alone. Science makes the truth about the world around us comprehensible to our puny human headmeats by means of a divide-and-conquer strategy (Cartwright 1999, Dupre 1995), and the use of differential algorithmicisations exemplifies this.

REFERENCES

Cartwright N. 1999. *The Dappled World*, Cambridge: Cambridge University Press.

Clark A. and Chalmers D. 1998. “The Extended Mind”, *Analysis*, 58:7-19.

Crick F H C. 1958. “The biological replication of macromolecules”. *Symposia of the Society for Experimental Biology* 12:138-163.

Dennett D. 1995. *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. London: Penguin Books.

Dupre J. 1995. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass.: Harvard University Press.

Field H. 1989. *Realism, Mathematics, and Modality*. Oxford: Blackwell.

Goldstein M H., King A P, and West M J. 2003. “Social interaction shapes babbling: Testing parallels between birdsong and speech”, Proceedings of the National Academy of Sciences, 100:8030-35.

Gould S J. 1997a, “Darwinian Fundamentalism”. *New York Review of Books*, June 12, 1997.

Gould S J. 1997b. “Evolution: the Pleasures of Pluralism”. *New York Review of Books*, June 26, 1997.

Gould S J. and Lewontin R. 1979. “The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme”. *Proceedings of the Royal Society of London, Series B* 205:581-598.

Koch K. 1999. *The Biophysics of Computation: Information Processing in Single Neurons*. Oxford: Oxford University Press.

Levelt W. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

Lewis D. 1991. *Parts of Classes*. Oxford: Blackwell.

Lotka A. J. 1925. *Elements of physical biology*. Baltimore: Williams & Wilkins Co.

Oudeyer P. 2006. *Self-Organization in the Evolution of Speech*. Oxford: Oxford University Press

Tooby J. and Cosmides L. 1997. “Letter to the Editor of The New York Review of Books on Stephen Jay Gould's “Darwinian Fundamentalism” (June 12, 1997) and “Evolution: The Pleasures of Pluralism" (June 26, 1997)", available at http://cogweb.ucla.edu/Debate/CEP_Gould.html

Vihman M M. 1996. *Phonological Development: The Origins of Language in the Child.* Oxford: Blackwell.

Volterra V. 1926. "Fluctuations in the abundance of a species considered mathematically". *Nature* 118:558-560.