Winner of the 2015 Boethius Prize

A Taxonomy of Causal Reasoning

by James Edwards

By Ratiocination, I mean computation. – Thomas Hobbes

This essay presents a philosophical tool, developed during professional research into fraud detection. The tool is a taxonomy that demarcates types of causal reasoning based on the objective or goal of the inquiry or assertion. It may be used to track the reasoning and arguments of others as to matters of fact, and to identify the necessarily implied occurrence of related evidence. The essay also seeks to identify the ultimate source from whence the instinctive belief in causation, understood in a Humean sense, arises. In addition, the underlying model, presented as a structural problem space, claims to be an architecture genuinely embodied in the human brain.

	← The Problem Space →					
	A	Δ	Ω	Mode/ Method/Type	Direction	Outcome
	(α_n)	**$f(\alpha_n)$**	**(α_{n+1})**			
1	Known	Known	Known	– Perfect Information –		
2	Observation	Theory	?	Deduction	→	Prediction
3	Observation	?	Observation	Induction	→←	Dicta
4	Observation	?	?	– Insoluble –		
4a	Observation	Conjecture	?	Eduction	→	Prediction
4b	Observation	?	Speculation	Anteroduction	→ (→←)	Dicta
6a	?	Theory	?	– Contingent –		
	(α_n)	**$f^{-1}(\alpha_{n+1})$**	**(α_{n+1})**			
5	?	Theory	Observation	Reduction	←	Postdiction
6	?	Theory	?	– Contingent –		
7	?	?	Observation	– Insoluble –		
7a	?	Conjecture	Observation	Abduction	←	Postdiction
7b	Speculation	?	Observation	Retroduction	← (→←)	Dicta
8	?	?	?	– Total Ignorance –		

Causal reasoning, according to David Hume (1777: 25-27), is reasoning related to *matters of fact* (as opposed to *relations of ideas*), and is “founded on the relation of *cause and effect*.” The notion of this relation, however, lacks intuitive or demonstrative certainty “discoverable by the mere operation of thought, without dependence on what is anywhere existent,” and is therefore never more than simple supposition; namely, that “there is a connection between the present fact and that which is inferred from it.” That such suppositions are, nevertheless, "of the human make and fabric," Hume attributes to a customary propensity *to believe*; a binding of fact to inference by way of *feeling*; the result of *habit* or *instinct*:

> "As nature has taught us the use of our limbs, without giving us the knowledge of the muscles and nerves, by which they are actuated; so has she implanted in us an instinct, which carries forward the correspondent course to that which she has established among external objects; though we are ignorant of those powers and forces, on which this regular course and succession of objects totally depends." (1777: 55)

Hume further recognizes that these customary suppositions, "implanted in us as an instinct," are critical to what we can, using our more modern parlance, plainly call human *problem-solving*:

> "Custom is … so necessary to the subsistence of our species, and the regulation of our conduct, in every circumstance and occurrence of human life. Had not the presence of an object instantly excited the idea of those objects, commonly conjoined with it, all our knowledge must have been limited to the narrow sphere of our memory and senses; and we would never have been able to adjust means to ends, or employ our natural powers, either to the producing of good, or avoiding of evil." (1777: 55)

It is prudent to grant Hume what is, after all, primarily an epistemic claim, and take heed of his caution that such an instinct is "in some cases apt to lead us into errors" and that "a wise man, therefore, proportions his belief to the evidence." (1777: 110) Nevertheless, Hume's very explication as to the source of the instinct contains, perhaps unavoidably so, the implicit ontological claim of ascribing a "regular course and succession" to nature, which is wholly distinct from the claim that man makes causal inferences instinctively and contrary to reason. Such an ontological claim as to the nature of reality has been labeled the *uniformity principle*, alternatively stated as: "the future will be conformable to the past." (1777:35) Let us now consider from whence such an instinctive belief in the uniformity of nature arises. Might we, perhaps, discover an ontological basis for such an instinct?

A Brief Evolutionary Foray

Evolution, the idea that those novel variations in an organisms' extant inheritable traits which provide a better environmental 'fit' (an 'adaptation'), via improvement in survivability or reproduction, are naturally selected, and thereby spread through a population over successive generations, offers us one very important, but often overlooked, lesson: namely, *scale*. While Hume lived approximately 250 years ago, and Aristotle approximately 2500 years ago, the *Mammalia* clade, with its distinctive neocortical brain structure, and from which our *Hominidae* family and *Homo* species evolved, first appeared approximately 250 *million* years ago!

It seems reasonable to suppose that *the distinctive brain* of whatever prototypical organism marked our evolutionary progression, likewise progressed in some environmentally adaptive fashion. We may, I believe, further assume that each such incremental adaptation brought along with it various tensions which, over time, sorted themselves out by some structural re-arrangements required to maintain the adaptation in an efficient manner.

I labor in these speculations only to emphasize that for hundreds of millions of years some evolutionary predecessor to man was employing something more closely akin to visuospatial representation, than the higher level abstract, propositional cognitive abilities we enjoy today. It is not that I contend visuospatial representation, or the like, underlies all causal reasoning, but rather that the organ (i.e., the brain) that we now use for more advanced forms of abstract representational reasoning has been exaptated, or co-opted, to accommodate these traits. In this view, the new adaptations are superimposed on previous adaptations, the organ exhibits characteristics of each, and ‘reasoning’ is perhaps more properly conceived as being on some continuous spectrum within a multi-level sensory-perceptual-mind *system*. If so, early activity along this spectrum, which again, lasted for hundreds of millions of years, was localized, immediate, tactile, and action-oriented.

Shastri & Ajjanagadde (1993) recognize a distinction between *reflexive* reasoning (that which is rapid, spontaneous, and without conscious effort) and *reflective* reasoning (that requiring reflection, and conscious deliberation). Such a distinction alone, however, while insufficient to separate human reasoning from that of other animals, betrays an earlier system of representation, upon which our current higher-order (more abstract) representational systems have been constructed. Darwin (1936: 453-455), for example, held that animals, who may be seen to “pause, deliberate, and resolve,” possess some power of reasoning. He cites both an elephant capable of bringing an otherwise out-of-reach object within direct reach by using his trunk to blow air just beyond it (so as to move it closer), and a restrained bear gaining an otherwise out-of-reach piece of floating bread by pawing the water (so as to create a drawing current), as examples of rudimentary, non-propositional, reasoning. Such acts are, Darwin claims, based on some learned association of coincidences, related in some way to searching, and “whether or not any general proposition on the subject is placed before the mind” are nevertheless “guided by a rude process of reasoning, as surely as a philosopher in his longest chain of deductions.”

Anderson (1990: *x*) achieved a breakthrough upon his realization that: “… human cognition could be understood on the assumption that it was an optimal solution to the information-processing problems posed by the environment.” Environments, however, are variable. If we seek to discover an ontological source for the human instinct that explains how the power of the Humean uniformity principle is generated, we need to go beyond variable environments. Why don’t we try what philosophers do best? That is, let's make that *bold categorical move* that philosophers are so wont to do. After all, aren't any and all environments merely instantiations of, well, um … *reality*?

A Brief Metaphysical Foray

Let's cobble together a simple metaphysics of reality. From Spinoza we shall adopt the notion that reality is but one thing (*substance*). From Democritus, the notion of atomism; which we shall tweak so as to require only *some basic unit* of substance, all sharing identical properties; whether the units be some fundamental nuclear particle, quanta of energy, or whatever. Lastly, let's add the notion attributed to Heraclitus that *everything flows*. If we contemplate such a reality, we might obtain the conception of One *system,* unified through its basic unit of substance, yet simultaneously capable of exhibiting variation due to local densities and relational configurations of the flowing units.

In such an ontology, the source of the uniformity is apparent: it is within the oneness of the system. That is, despite the unit flows and local configurational variations, the basic units, their inherent properties, and their infinitude or absolute number, do not change. Thus we have a single system whose state changes, as determined by arrangement of its constituent units.

If we imagine any instant of unit flow to be infinitesimally small, a mathematical structure which would convey such oneness with configurational variation would be a sequence. Let's consider the basic mathematical sequence:

$$(\ldots\ 0, 1, 2, 3, 4, 5\ \ldots)$$

Such a sequence, while indeed conveying a regular progression of oneness, is nevertheless suspect. The flow from state to state seems determinable by no more than serial position within the sequence. This won't do.

Russell (1919: 199), in considering a *closed system* (as our single reality should arguably be treated), modifies the basic sequence by adding a functional relation based on time, such that:

$$(\ldots\ E_1, E_2, E_3, E_4, E_5\ \ldots\ E_t\ \ldots)$$

where $E_t = f(e_1, t_1, e_2, t_2, \ldots, e_n, t_n, t)$ and e_n and t_n are state data and times, respectively, in a closed system; for which, in terms of our single closed system of reality, $e_1 = E_1$. Now while an apparent improvement has been made, namely, the flow has been functionally accounted for, it has only been accounted for as a function of time; and this seems unacceptable. It is as if we introduced time merely as an accounting convenience, so to speak, and not because time was in any way intrinsic to the system. Russel himself seems to notice this:

> "All mechanical laws exhibit acceleration as a function of configuration, not of configuration and time jointly; and this

> principle of the irrelevance of time may be extended to all scientific laws. In fact we might interpret the "uniformity of nature" as meaning just this, that no scientific law involves time as an argument, unless, of course, it is given in integrated form, in which case lapse of time, though not absolute time, may appear in our formulae. Whether this consideration suffices to overcome our difficulty completely, I do not know; but in any case it does much to diminish it." (Russell: 205)

An alternative sequence, which avoids the artificial introduction of time, is the recursive sequence. A recursive sequence is one where any term is defined in terms of some earlier term, which for our purposes we may define as the instantaneously preceding prior term. So, while the system sequence may look identical to that of Russell's construction:

$$(\ldots E_1, E_2, E_3, E_4, E_5 \ldots)$$

the underlying formulation is different; namely $E_{n+1} = f(E_n)$; or, alternatively $f(\alpha_n) = (\alpha_{n+1})$. The recursive sequence comports to our dual requirement: namely, the system states retain their oneness, while the transitions between states are accounted for by no more than the previous state. The function is *one-to-one*, or bijective, which maps every (α_n) to one and only one (α_{n+1}), and vice versa. Thus, provided *the inverse of the function* is used as appropriate, the equation is wholly reversible. Were state changes in a recursive sequence to be calculated mathematically, *difference* equations instead of differential equations would be required.

We are prepared, at last, to establish how it is that reality itself generates the instinctual belief in its uniform regularity. That we rely upon mathematics to do so is fitting, for as Galileo states: "This book [of the universe] is written in the mathematical language."

The Problem Space

Johnson-Laird (2004: 179) cites Kenneth Craik as progenitor of the concept of *mental models*:

> "If the organism carries a 'small-scale model' of external reality and its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past event in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to emergencies which face it." (Craik, as cited in Johnson-Laird:2004).

As part of their more general *Theory of Human Problem Solving*, Newell, Shaw and Simon (1958) and Simon and Newell (1971) put forth the notion of *problem-space*, which they define as: "the way a particular subject represents the task in order to work on it." (1970: 151). While recognizing and adopting the centrality of the problem-space notion, our use differs from that of the original authors. Newell, et. al., as well as many more modern scholars, seem to adhere to a *modularity of mind* conception that invokes multiple, task-specific modules. Some even go so far as to suggest *massive modularity*. Our conception of the problem space, on the other hand, is one of a single, consistent, general purpose architecture that operates between Newell's information processing *biological* and *cognitive bands*, where it is accessible to both automatic and controlled processing.

This architecture has a number of interesting aspects, of which two are quite noteworthy. First, it is conceived as a series of data storage nodes or *registers*. Unlike short-term memory *per se*, a register holds data *in-processing*. Each register seemingly has preference for data of a certain *type* or structure. Second, the three nodes are arranged, or ordered, *computationally*, in the form of an algebraic equation. Our first node (α_n) is part of the computational set, but is only *calculated* as part of the $ƒ(\alpha_n)$ transformation. The (α_n) node seems to show a sort of preference for data which represents the knowledge state of the observational field, although the contents of the Alpha node (α_n) can be transposed with contents of the Omega (expectant) node (α_{n+1}), so as to facilitate reasoning in imaginary predecessor or successive problem spaces.

Ordering as a computationally arranged algebraic equation allows the nodes to be operated upon by various independent subprocesses which perform the algebraic equation-solving tasks of substitution, re-arrangement, and elimination which are necessary to obtain *the solution for the unknown quantity*. The problem space 'forces' an algebraic computational structure upon the task for reflexive reasoning, with some variable degree of penetrability, and is open, accessible, and 'visible' for reflective reasoning. This is in accord with Holyoak & Spellman (1993: 276), who suggest that "reflexive reason of the sort involved in ordinary comprehensive relies on dynamic binding of objects to argument slots in preexisting rules." Further, as Jevons (1869: 16-23) points out, "we always reason by means of identities or equalities" and "… the true secret … [is] that when there is no equation, no inference can be drawn." By setting $ƒ(\alpha_n) = (\alpha_{n+1})$, the mind adheres to Jevons' principle that "all forms of reasoning consist in repeated employment of the universal process of … *substitution of similars*."

That our algebraic equation (of reality) is itself identical to Newell's *representation law* (entity [X] is transformed [T] into entity [Y]), is the reason why the uniformity principle is maintained via our causal instinct. It is a strict data content to data content recursive correlation which preserves both ontological oneness and transformational variability; ensuring that *exactly* similar causes will produce *exactly* similar effects. This alone does not, however, answer our question as to where the *power* of the belief comes from. To my mind, this power springs from the underlying mathematical equation. That is, there is something about an empty slot in a computational algorithm that compels the insertion of *some* value; and, once inserted, equally compels an inference to the unknown quantity. As Newell (1992: 430) states: “architecture creates its own goals whenever it does not have what is needed to proceed.”

Taxonomical Observations

The contents of each node represents knowledge states as to the data contained in that node. Nodes were filled-in according to the fashion of completing a truth table. Two ‘knowns’ are required for a solution. *Speculation* and *Conjecture* are downgraded states of *Observation* and *Theory*, respectively. The ‘?’ is the target goal (*the unknown quantity*), which can be achieved either in action or simulatively.

The taxonomy indicates that there are three primary modes of causal reasoning (deduction, induction, and reduction), with four derivative modes, for a total of only seven possible modes.

By providing a stable and reliable interface to the world, the problem space goes a long way toward ensuring successful problem solving. The subprocess of system decomposition, or what Russel (1919: 198) calls “practically isolated” and what Simon (1962: 474) says is founded upon a *theory of nearly-decomposable systems* remains. In short, careful decomposition requires precise *configuration specifications* for (α_n) or (α_{n+1}) objects, and a precise *particularized generalization* for $f(\alpha_{n+1})$.

The underlying problem space model suggests why logicians such as Jevons (1874: 14) hold deduction to be the preeminent mode of causal reasoning (“… all reasoning is founded on the principles of deduction”). While all modes allow for computation, only the deduction mode allows for direct forward calculation. So-called *sense-making* (plausibility) seems associated with the deductive mode as well.

While it is theoretically possible to reason directly in modes requiring backward reasoning, few seem capable of routinely implementing the required inverse functional transformation; and therefore it seems that these problems are computed backward, so to speak, but calculated forward on an iterative construct–test basis.

That conditionals often prove difficult may be explained by the taxonomical interpretation that the conditional transformation is implicit. Further, the fallacy of affirming the consequent may be seen as a failure to note the necessity of inverting the transformative function.

It was, perhaps, not so much that algebra was ‘invented’ in ancient Babylon, but that the very same logic espoused herein was introspected, distilled, and codified.

References

Anderson, J.R. (1990). *The Adaptive Character of Thought*, Erlbaum.
Darwin, D. (1936) *The Origin of Species and The Descent of Man*, Modern Library.

Holyoak, K.J. & Spellman, B.A. (1993). Thinking. Annual Review of Psychology 44, 265-315.

Hume, D. (1777). *Enquiries Concerning Human Understanding*, Selby-Bigge, L.A. (ed.); Oxford University Press (1975).

Jevons, W.S. (1869). *Substitution of Similars*. MacMillan & Co.

Jevons, W.S. (1874). *Principles of Science*. MacMillan & Co.

Johnson-Laird. P.N., (2004) “The history of mental models,” in Manktelow, K. and Chung, M.C. (eds.) Psychology of Reasoning, Psychology Press.

Newell, A. (1992). Précis of Unified Theories of Cognition. Behavioral and Brain Sciences 15. 425-492.

Newell, A., Shaw, J.C., and Simon, H.A. (1958). Elements of a theory of human problem solving. Psychological Review 65-3, 151-166.

Russell, B. (1919). “On the notion of cause,” *Mysticism and Logic*, Longmans.

Shastri, L. & Ajjanagadde. V. (1993). From simple associations to systematic reasoning. Behavioral and Brain Sciences 16-3, 417-494.

Simon, H.A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society 106-6, 467-482.

Simon, H.A. and Newell, A. (1971). Human Problem Solving. American Psychologist 26-2, 145-159.