Causality in Cognition: Do we need a Fourth Contender?

Philosophical Issues in Cognitive Science

Professor Anthony Corones

Semester 2, 2005

Okko Buss, 3152903

The problem of accounting for causality has riddled all attempts at explaining cognition. Roughly stated, the problem lies in explaining how internal states cause action or other internal states. It thus underlies a rich tradition of philosophical debate from dualism to materialism, the demise of behaviorism and still goes remains unaccounted for in most modern disciplines of Cognitive Science.

Three modern theories of cognition (and more or less loosely attached Artificial Intelligence research programs) compete in offering suitable accounts for causality: Folk Psychology (e.g. Fodor, Pylyshyn, Dennett), Connectionism (e.g. Paul and Patricia Churchland, Sejnowski, Rumelhard) and most recently, Dynamic Systems Theory (Eliasmith, van Gelder, Clark). None have had resounding success, though each have found some, if different, application to a range of mentalistic events that can, broadly be classified as causal. This paper seeks to provide an overview how causality is treated in these approaches and demonstrate that this difference in scope calls for a more unified or more open-minded approach, or, to co-opt Chris Eliasmith's phrase, a "fourth contender". Such a contender may be a taxonomy of mechanisms found in intelligence, as proposed by Aaron Sloman.

Folk Psychological explanations, most notably in the vein of Jerry Fodor's Language of Thought rely on the causal power of propositional attitudes, such as beliefs and desires, to account for the cause of action in cognizers. For example two such propositional attitudes, the belief that 'dark clouds mean that it will rain' and the desire to 'stay dry and warm', have the causal power for actions, such as to make one seek shelter. Falling back on propositional attitudes is thus sufficient to account for intelligent behavior.

In this account of cognition, causality is thus taken as a primitive, i.e. taken for granted, as a intrinsic aspect of propositional attitudes. Believing that it will rain and not wanting to get wet is a sufficient account of how behavior arises as a result of

some combination of internal and external states. This causal power is labeled "intentionality" as in Daniel Dennett's "intentional stance" (Dennet, 1987), which, in contrast to his "physical" and a "design" stances, captures this level of explanation. In computational theory and AI, this stance is roughly analogous to the "symbol" level, as in Pylyshyn's description of classical computational architecture (Pylyshyn, page 58). At the very least, both make reference to semantic constituents that are said to be rule-governed (i.e. 'symbols' in symbolic AI and 'attitudes' in Folk Psychology), but in terms of the issues concerned here, both attribute intrinsic 'intentional', causal powers to these constituents.

Intentionality takes the properties of an intelligent agent at face value. Such a agents are, as Andy Clark pointed out "semantically transparent systems (Clark, page 29)". Symbols and their manipulation, propositional attitudes and their 'intentions' are thus read directly off the system and given a "familiar" interpretation. However the assumption that this is possible may be a deep illusion about the nature of cognition. Intentionality, in fact runs into several problems when confronted with the task of accounting for causation. Four seem significant.

Firstly, if a propositional attitude, such as a desire (e.g. to stay dry in the rain) is said to have causal powers (e.g. make the desirer go seek shelter), can it reliably be shown that this attitude was in place at the time of action? There is a distinct possibility that it was not, rather that the 'attitude' is an *a posteriori* justification or explanation of behavior. This effect has occasionally been demonstrated in psychological experiments on consumer decision making[1] and suggestion under hypnotism, in which the hypnotized develop reasons for why they followed otherwise irrational orders. While it is not clear whether this criticism applies to all types of propositional attitudes, nor what precisely 'being in place at time of action' really

1 E.g. http://srsc.ulb.ac.be/axcWWW/papers/pdf/01-IESBS.pdf

means, in terms of propositional attitudes, it draws into question whether recourse to the causal powers of such attitudes is truly "necessary and sufficient (Newell and Simon, page 87)".

Secondly (and relatedly) there seems to be a collapse of "causation" and "reason". Causation as a mechanical or physical concept is relatively well-understood. However it is far from clear that this understanding translates aptly when dealing with mentalistic discourse (i.e. the Language of Thought). Folk Psychology is thus faced with the double burden of justifying the assumption that there are attitudes prior to action or behavior, as well as demonstrating that such attitudes are strictly "causal", rather than "reasons" that aren't the same as (rather than, perhaps, symptoms of) mechanisms with true "causal" power.

Thirdly, there are non-propositional aspects of cognition, which require a better account for causality than Folk Psychology can provide. For example, motor control, music perception and speech production (three aspects of cognition that Connectionism and Dynamic Systems Theory have had success in modeling, see below) defy propositional explanation. For example, experiments in motor control (the "two visual systems" experiment of Aglioti, Goodale and Desouze, 1995) show that grabbing an object according to its correct size does not depend on the conscious perception of it, which can be subject to a visual illusion. Whatever account Folk Psychology gives, it cannot explain propositionally how such correct behavior is caused. Even "simple", everyday tasks such as walking occur largely without formulation of propositional attitudes that hold causal power to drive such tasks.

Lastly, there is a logical problem with attributing causal powers to propositional attitudes. If in some theory of mind, "causality" is defined as the power of such propositions, but these propositions are the very target of that theory, nothing has really been achieved. This definition risks being either trivial ("Q: What is causality?

A: Everything that is necessary for intelligence") or circular ("A: Propositions with causal powers"), which according to Aaron Sloman is generally true of purely "computational" theories (Sloman, 1996). Folk Psychology is notorious for this in it's account of causality in intelligence.

In summary, Folk Psychology perhaps offers some explanatory insight at levels of mentality where propositional attitudes can be shown to exist. However it succumbs to the illusion of semantic transparency of cognizers, glosses over causes of mental events that require non-propositional accounts and generally suffers from triviality and circularity in its argumentation. As such, it merely shifts the focus of the question "how is causality achieved in cognition?" to "how is causality achieved by propositional attitudes?" without adding any original explanatory oomph.

The second of our three paradigms, Connectionism, draws on this particular shortcoming of classical, symbolic A.I. (Folk Psychology's cousin) to overcome the cognitive illusion of semantic transparency. Classical A.I.'s successes have become fewer and fewer, even as computational speed and power are rising. In proposing more brain-like computation and by representing a combined A.I. toolkit and philosophical/conceptual framework under one roof, Connectionism rise in the 1980s brought with it several significant improvements: computation became 'parallel' and 'distributed', i.e. generally more brain-like, which served to disperse almost instantly the illusion of semantic transparency that riddled Folk Psychology. Causality in a Connectionist framework is no longer assumed as a primitive, but given a mechanistically embedded definition. Causality is what the brain does.

This focus on mechanism (i.e. modeling the neuronal properties of the brain rather than its functions) allows for reliable input-output mapping in connectionistic models, without falling back on flawed notions of semantic transparency. This is especially true for those aspects of cognition that fall outside the descriptive range of

"sentential epistemology", as Paul Churchland termed Folk Psychology. Good examples of these that have been demonstrated on connectionistic software, neural nets, are vision and motor-control (Churchland, 1992), speech production (Sejnowski, 1987) and music perception (Leman, 1995). NETtalk (Sejnowski, 1987) represents a learning system that acquires the ability to produce English phonemes from English text without explicit instruction. Leman (1995) trained a neural net that had the ability to recognize tone-centers and musical harmonies, only from listening to 300 iterations of clean harmonies. It is important to realize that although there's a descriptive theoretical canon for these abilities, not directly concerned with cognition, (e.g. Phonology, Musicology), they are generally semantically opaque[2].

Connectionism's success in modeling non-propositional behavior, without succumbing to the flawed assumptions of Folk Psychology, raises two questions about the range and types of phenomena each offers explanations for:

1) Do the semantically transparent aspects of cognition that Folk Psychology attempts to describe also fall under the blanket of semantic opacity, as already suggested by psychological and hypnotic experimentation (see above).

2) Can connectionist models also apply to such ostensibly transparent, but likely opaque processes (or if they are simply out of scope, what does that say about the illusion of semantic transparency?)

Unless it turns out that the questions of Folk Psychology about our beliefs and desires and their causal powers can be cleverly rephrased, most certainly the true test for Connectionism (and the ultimate demise of Folk Psychology as anything more than folk) will lie in answering the second question. Unfortunately it can only be answered theoretically today, as neural nets continue to fail to impress with displaying

2Linguistic literature for speech production uses highly symbolic language of phonology, even if 'symbolic' here has a different meaning from the one used to describe Folk Psychology. Interestingly though, this may still be the reason why NETtalk still requires a Text-To-Speech (TTS) front-end, a hallmark accomplishment of symbolic A.I.

truly remarkable intelligence (even at higher computational power).

At the very least, by providing a quasi-mechanistical model for an albeit limited range of cognitive phenomena, connectionism treats causality no longer as tacitly assumed, but as directly embedded in the system itself. However for such phenomena falling under "sentential epistemology", connectionism has yet to provide a working model. The debate whether a neurological brain-like account for more abstract mentalistic events than producing phonemes or perceiving & distinguishing musical harmonies is possible remains unanswered[3].

The "third contender (Eliasmith, 1998)" is Dynamic Systems Theory. This approach employs the language and tools of mathematical dynamics (especially differential equations), and thus represents a "novel set of metaphors (Eliasmith, page 306)" purportedly with distinct advantages over both Folk Psychology and Connectionism. Indeed it is immediately attractive for sharing features with both Folk Psychology and Connectionism. It resembles Folk Psychology by allowing for levels of descriptive abstraction not accessible in Connectionism (whose reductionism to neurology makes it restrictive). And like Connectionism, Dynamic Systems Theory does not succumb to the flawed assumption of semantic transparency.

Just briefly, the technical language of Dynamic Systems Theory refers to descriptions that are "deterministic", "non-linear", "relative to time", "low-dimensional" and "coupled". The first three seem to underlie the similarity with Connectionism. The fourth, "low dimensionality", represents a similarity with Folk Psychology, which "is a feature which contrasts the dynamicist approach with that of the connectionist (Eliasmith, page 308)", also indicating that the conceptual distance from Dynamic Systems Theory to Folk Psychology is far greater than that to Connectionism. The last, "coupling", however represents an original concept that

3Connectionist seem to skirt the issue by declaring only "sub-conceptual" answers matter.

directly concerns how causality is accounted for. This part of the novel metaphor sets it apart from either of the first two contenders.

Coupling describes the relationship of two functions or mechanisms (again captured by differential equations) in which the input of each is the output of the other. Two such mechanisms are thus causally coupled, which, in the language of Dynamic Systems Theory, is to say "reciprocal causality". The favorite example of this in the literature is that of a Watt engine driving a flywheel, which requires a constant speed, achieved by regulating the flow of steam through a steam valve (van Gelder, page 347f). A 'computational' implementation of this regulation involves multi step instructions from measuring the speed of the flywheel comparing it to a desired speed and then adjusting the throttle valve. A more dynamicist implementation was eventually favored by Industrial Revolution Engineers, who included a spindle with arms that move upwards at higher speeds (i.e. when the throttle valve opens) in turn closing the throttle valve. The two mechanisms, spindle and valve, are thus continuously reciprocally causal, or coupled. This coupled relationship is mathematically captured, thus describing the causal interaction between the two.

Dynamicist examples of this type from cognition are rare. Moreover, the most convincing ones, as in Connectionism, tackle aspects of cognition of low abstraction, such as motor or sensory control, for example treadmill stepping in infants (Clark, page 124) or the olfactory bulb (Skarda and Freeman, 1987). This particular range of explanation clearly mirrors the fact that, in terms of accounting for causality, Dynamic Systems Theory also sets itself apart more from Folk Psychology than from Connectionism. In addition, the conceptual advantage of modeling a greater range of cognitive phenomena of any level of abstraction (so long as they are mathematically outlined) that Dynamics purports to have over Connectionism is thus not apparent.

This may in part be due to the difficulty of determining reciprocal causal effects between internal and external states (e.g. cognizer and environment) which are difficult to determine at the best of times. Whatever edge Connectionism has over Dynamics or vice versa in explaining causality thus remains to be resolved.

These three approaches then have applied explanatory models to various different aspects of cognition in the fashion outlined above. Crucially, they differ in terms of how they attribute causal powers to internal states, whether propositional, neuronal or dynamical and how those powers are said to give rise to other states and intelligent behavior. However while these differences in approach have yielded more or less satisfactory accounts of causality, very little has been revealed about the nature of causation in cognition in general. This mostly concerns Folk Psychology and Connectionism, in which causality finds a limited definition as either propositional primitive or patterns of neuronal activation, respectively. Dynamics, even if not explicitly committed to any type of causality so long as it is expressible in the language of differential equations, still has very little to say about it other than it's reciprocal, embedded nature.

A very different type of approach from these has been put forth by Aaron Sloman (1996), who calls for examining "the potential uses of all relevant mechanisms, whether computational in any sense or not. (Sloman, page 191)" in explaining cognition. Sloman proposes a type of taxonomy of mechanisms required for intelligence, employing conceptual distinctions along the lines of computation, neurology and physiology as part of this taxonomy rather than as guiding *a priori* principles, and thus allowing for a line of exploration unbound by conceptual prejudices.

Sloman begins by pointing out what he perceives as the crucial shortcoming of

a purely computational approach. Essentially, all computational definitions of intelligence tend to be circular or trivial, much in the same way that propositional attitudes suffer from circularity or triviality (or both) when accounting for causality (see above.) Computation to him is out as a unified account, because "it is not clear ... that we can find an alternative useful general definition that covers all the interesting cases and avoids the twin traps of circularity (because it presupposes some aspect of intelligence) and triviality (because it implies that all processes are computation.) (Sloman, page 184)"

Moreover, Sloman claims that Connectionism does not satisfactorily avoid this problem either since "if we allow 'computation' to include possibly non-discrete brain processes, slide rules, and soap bubbles then it is hard to see how we can draw a boundary between computations and non-computations. (Sloman, page 183)" While Sloman may be too happy to readily conflate Connectionism with symbolic computation in his reasoning, he does however acknowledge that "there is no *conceptual* reason to rule out essentially continuous mechanisms from playing a useful role in human-like mental processes." In light of his attempt to find "the potential uses of all relevant mechanism" this caution to avoid one-theory explanations seems entirely reasonable.

Accounting for mechanism alone, however, Sloman believes does not necessarily account for causality. His concern therefore extends to causality, or more specifically "what is control", which he deems "a special case of the notion of causation (Sloman, page 186)" captured in the causal relationship between controller and controlled. While this resembles the Dynamics notion of coupling, Sloman extends it to a range concepts that may or may not be captured in differential equations along the lines of Dynamic Systems Theory (such as 'Virtual Machines' or the interplay between compiled and interpreted programs in computer science).

Importantly, accounting for mechanism in the language of engineering (rather than mathematics or neurology), Sloman points out must also include an account of control and causality, an undertaking best commenced by taxonomy.

For instance, in computer science, a program's (or more aptly a Virtual Machine's) control over the processes it spawns can be described in a number of feature pairs, some of which are 'compiled' or 'interpreted', 'direct' or 'indirect', 'open-' or 'close-loop' (notions that mirror the Dynamic considerations of 'embeddedness' in the environment), 'ballistic' or 'online' (does the process retain control after causing another process or not?), and more or less 'virtual'. This last case, of 'virtual' processes, extends to the discussion of Virtual Machines, which Sloman points out are not in contrast with "real" processes, rather than "abstract" ones (Sloman, page 188). Virtual Machines represent a commitment to exploring all possible levels of explanation, and directly defies Connectionist reductionism. As Sloman is wont to point out, positing a single "ultimate" layer of explanatory reality (Sloman, page 188f) is probably inadequate. "This requires drastic rejection of most common-sense concepts of causation and control." Sloman's taxonomy of more or less virtual processes thus retains a powerful aspect Folk Psychology, that abstract processes may indeed be talked about.

In sum, Sloman engages in a taxonomy or intelligent architectures by retaining key notions from all other 'contenders'. From Folk Psychology he inherits such terminology as "belief-like and "desire-like" states, "propositions" and a generally abstract approach, from Connectionism he takes the notion of non-discreteness and continuous mechanisms, and with Dynamics he shares notions such as reciprocal causation of particular mechanisms and the important consideration of the environment. For instance, in a single passage Sloman suggests that "the ability to give an internal substate a supposition-like vs belief-like state depends on the causal

links with the environment" and that "biological evolution includes developments along the directions indicated here (Sloman, page 210)". The freedom to engage in discussion of "belief-like states", "causal links to the environment" and "biological evolution" hopefully will lead to a more unified approach to cognition, unconfined by conceptual considerations such as computation, neurology or a descriptive mathematical metaphor.

Sloman's work is a proposed research program and stays within bounds of few technologically and cognitively limited examples, such as thermostats. However by imbuing existing theory with an open-minded notion of what control in cognition, and thus causation, may be, it appears to be an entirely plausible and fruitful approach. Theoretical accounts cannot be judged and compared transitively (X's explanation is better than Y's is better than Z', therefore X's is better than Z). As seen, the three most common paradigms, contenders, display a considerably different scope of events and phenomena for which they offer explanation. By remaining open to all of them, an account such as Sloman's, asking "what are the potential mechanisms", rather than "how is the mind like a computer/just a brain/a steam engine" probably has the best chance of success. Whether Sloman's taxonomy of computer control can be meaningfully extended to control in cognition remains to be seen.

References

- Aglioti, S., Goodale, M. and Desouza, J. (1995) "Size contrast illusions deceive the eye but not the hand", *Current Biology*, 5, 679-685 (from Clark 2001)
- Churchland, Paul (1992) "The Computational Brain", Cambridge, MA, MIT Press
- Clark, Andy. (2001) "Mindware", Oxford Univsersity Press, New York, New York
- Dennet, Daniel. (1987) "The Intentional Stance", Cambridge, MA, MIT Press
- Eliasmith, Chris. (1998) "The Third Contender: A Critical Examination of the Dynamicist Theory of Cognition" in Thagard, Paul. [ed.] "Mind Readings", Cambridge, MA, MIT Press
- Leman, Marc (1995) Music and Schema Theory: Cognitive foundations of a Systematic Musicology, Springer Verlag, Berlin
- Newell, A. and Simon, H. (1976) "Computer science as empirical inquiry: Symbols and search.", *Communcations of the Association for Computing Machinery*, 19, 113-126 (from Clark, 2001)
- Pylyshyn, Zenon. (1991) "Computing in Cognitive Science", in Posner, M. [ed] "Foundations of Cognitive Science", Cambridge, MA, MIT Press
- Sejowski, Terrence. (1986) "NETtalk: A Parallel Netowork That Learns to Read Aloud." (Technical Report JHU/EEC-86/01.) Baltimore, MD: John Hopkins University (from Clark, 2001)
- Skarda, C.J., and Freeman, W.J. (1987) "How brains make chaos in order to make sense of the world.", *Behavioral and Brain Sciences*, 11, 1-23 (from Eliasmith, 1998)
- Sloman, Aaron. (1996) "Beyond Turing Equivalence" in Milican, P.J.R., Clark, A. [eds] "Machines and Thought: The Legacy of Alan Turing", Vol 1, Claredon
- van Gelder, Tim. (1995) "What Might Cognition Be, If Not Computation", *Journal of Philosophy*, 91, 7, July 1995, 345-381