Causes and clauses and the Model Six

Stephen Larson
Greg Detre
May 2003

The notion of "Causes and Clauses" is one of the most appealing theories in Minsky's The Society of Mind. In this paper, we seek to reconcile this abstract idea with some of the more structured theories from The Emotion Machine. In particular, we have asked ourselves, how can the notion of "Causes and Clauses" fit together with the notion of a six-level model of mind?

Let us remind ourselves briefly what these two theories say. "Causes and Clauses", described prominently in The Society of Mind, but also in an essay entitled "Alien Intelligence", is a description of tendencies we use to represent the world. The theory has four parts: Things, Differences, Causes, and Clauses. Things are components of a scene, roughly corresponding to nouns. Differences are a comparison of different things, and can themselves be Things. Causes are the person, process, or things that we hold responsible for differences. Clauses are structures that can be built out of Things and Differences, and that can describe Causes. Clauses can also be Things. One of the features of this theory is the way that all four parts can be considered special kinds of Things. The ability for Clauses to be considered Things allows Clauses to be parts of Clauses, and for chains of thought to be composed recursively.

The six-level model of mind (Model Six) is detailed in the fifth chapter of The Emotion Machine. It includes levels that go from most simple at the bottom to most complex at the top. At the bottom of the model are instinctive reactions such as reflexes and reactions that we have at birth. Learned reactions are the next level up, and are made up of those reactions that we acquire over time. Deliberative thinking, on the next level, describes forward-looking kinds of thought that can test and act out hypothetical plans. Further up, reflective thinking allows us to look back on the results of deliberative thought, recognize patterns, and improve our future deliberations. Further up yet, self-reflective thinking adds a model of self to the actions of the world, and considers that self's actions from an outside perspective. Finally, self-conscious thinking incorporates the perceived opinions of others into evaluations about what the self should be doing.

Before moving into an attempt to reconcile the two models, let us first motivate the use of a theory of mind which endeavors to decompose the world down into constituent features as a strategy for understanding it.

Intelligence As Decomposition

There is a knockdown argument, used by Ned Block to attack the validity of the Turing test as an infallible indicator of intelligence that we initially worried might be applicable here. It might go something as follows: we might imagine an alien species with a very, very large head, that doesn't do any learning whatsoever, but is born with a huge list of rules for what to do in any of the situations it will ever encounter. We can see this as a huge instinctive-level only brain. Within the terms of the argument, such an alien would be intelligent; at least by any behavioral test we can devise on that world, but would have no need to decompose its representations. This need not worry us though, for two reasons.

Firstly, the combinatorial explosion involved in passing even a five-minute Turing test reliably and administered imaginatively using just a lookup table would almost certainly exceed the storage capacity of a computer the size of a planet. In fairness, Block's argument is intended to be an *a priori* one, and therefore not relevant to us, given that we're only concerned with what is physically possible. No alien with just a massive lookup-table for a brain could actually exist.

Secondly, such an alien would never learn. If its environment was to change in even the slightest way from that for which it had evolved, its rules wouldn't work. In fact, for the lookup table to work even in a static environment, that environment must be small enough to enumerate all the possible states and associated actions. For if the stimuli to which the system had been programmed to respond were features of the environment, or some kind of rules, then both of these would be kinds of decompositions that could not be represented on a lookup table. Even if we employ some sort of statistical classification algorithm like a self-organizing map to deal with new but similar situations, it would still be decomposing the situation into features of some description. The fact that we term the features 'sub-symbolic' seems to make them somehow differentiable from real, symbolic features. But this is a mistake – all 'sub-symbolic' means is that the features being detected are at a lower level or divide up the space in a different way from the features that we happen to be used to. Any sort of prediction requires figuring out what's similar about the present case to the past case. Unless they're identical, this means paying closer attention to some features at the expense of others. This is decomposition.

We feel that we can argue with some confidence that any intelligence, no matter how alien, must decompose the world into interacting components. Decomposition into symbols is important and possible because the world is structured and regular. This has the dual effect of allowing the system to ignore uninteresting features/components when making comparisons, and to be able to deal with many more (a potentially infinite number) of situations than it could possibly represent in atomic, list form. It is worth saying that there is, of course, considerably more to intelligent behavior than simply decomposing the world in some useful, predictive way. Firstly, you want to decompose it in many different ways, according to need and circumstance. Secondly, many of the hard, 'creative' aspects of intelligence are about *re*composing symbols on the fly in new ways.

Now that we have established what the two theories say, and justified the use of a strategy which decomposes perception into smaller bits, let us take a stab at integrating the two theories. Our strategy will be to first form a naive marriage between the two, examine where it falls short, and then attempt to combine them in a more profitable way in light of those shortcomings.

A Naive Combination

The first combination we will examine is motivated by the view that the Causes and Clauses idea describes a hierarchy. Since Model Six is clearly a hierarchy, a first attempt at unifying the two might try to match the hierarchical levels together. Let us explain why this might seem like a good idea at first.

The hierarchical structure of Causes and Clauses is not too difficult to see. Clauses are non-existent without being composed of nouns and verbs, to which Things and Differences relate. One could certainly describe a Cause in a Clause. It seems like Clause is in several ways the "highest-level" of this theory. Causes cannot be identified without recognizing Differences. Differences are only meaningful as a comparison between Things. Thus there appears to be a hierarchy of compositionality as one progresses from Things to Clauses.

How can we map this hierarchy onto the hierarchy in Model Six? It appears as though one can simply associate the Causes and Clauses idea with the first four levels of Model Six. It is easy to think about instinctive reactions as involving sensory perception. Minsky describes instinctive reactions as stimulus-response activities. For the purposes of matching these hierarchies, let us therefore think about Things as sensory stimuli. The first level of understanding of the world that a baby perceives might be described as a world full of stimuli that it does not yet understand, but responds to in ways that are genetically pre-programmed. In this way, as we are trying to combine these two hierarchies from the bottom up as we might zip up a zipper, we notice that this combined hierarchy seems more like a hierarchy of child cognitive development than a hierarchy of compositionality.

As we continue to zip up these two hierarchies into one, we find ourselves comparing Differences with Learned Reactions. As the developing child begins to make sense of the world, she starts categorizing stimuli by their similarity or difference. Only by a process of dissociating stimuli and identifying the differences can any learning occur. Thus Differences appear to be crucial to this level of Model Six.

Further upwards we compare Causes with deliberative thinking. Keeping the child development example in mind, it seems that only once a child has the ability to begin assigning agents of action to the differences she is observing can she begin to mentally experiment with deliberation. Because deliberation involves thinking about action in the future--action which is set in motion by an agent--the notion of Cause appears to be an important precondition. Children certainly cannot begin to identify causes for things before they are able to group stimuli based on their similarity or learn differences. It seems that Causes are appropriately matched with this level of Model Six.

Finally, we compare reflective thinking with Clauses. The fit between these two levels is rather straightforward. Because a clause allows the composition of the other elements, including itself, it is a structure well suited to the process of reflection. The clause can be thought of as the Petri dish or test tube of thinking, a container for fragments of ideas that allow them to be manipulated and examined as separate units. And since they Clauses are themselves things, they further lend themselves to reflection as they can be analyzed and composed in a recursive manner.

Better Combinations

This thought experiment, though interesting, has its limits. The most striking problem with the naïve combination is that prevents anything that is very interesting from happening within each of the levels of Model Six. Because Model Six is intended to be a description of processes layered on top of one another, rather than an ontology of representations layered on top of one another, we have a fundamental mismatch when we try to superimpose the two. What is needed, therefore, is a way to use the Causes and Clauses representations to construct ideas of how each level might operate; thus creating an explanation of Model Six in terms of processes rather than representations. We believe that a more profitable way to think about these ideas is to examine each level in turn and to determine a role for the parts of the Causes and Clauses framework. If a level does not appear to have a part, we make note of why it does not fit. If a level appears to require a part that does not exist, we identify where the framework needs to be improved.

Instinctive Reactions

Let us begin with the Instinctive level and take the example of a person who, by reflex, pulls their hand away after touching a hot stove. We can think of two different types of Things that apply to this level. Type 1 Things are low-level perceptual data, roughly represented as a feature vector. Type 2 Things are low-level motor responses. A Difference can be framed as the informational “distance” between perceptual data or motor responses that arise from different stimuli. In this case, the type 1 Things are the percepts of a regular hand and a burning hand, and the Difference is a measure of the way that those two percepts are dissimilar. The type 2 Things are the reflex motions that pull the hand away from the pain. Our notion of dissimilarity is inspired by the dot product between two vectors, which yields a numerical measure of similarity by quantifying how much the vectors point in the same direction. Our notion of dissimilarity can be thought of as the opposite of this measure.

When we attempt to discover places for either Causes or Clauses at this level, we turn up empty-handed. It doesn’t make sense to have a Cause for an instinctive reaction because the association between stimulus and response (type 1 and type 2 things) cannot be attributed to anything other than the genetically predetermined association between pain and reflex. While we might say that there was a cause for the reflex when looking from outside the system, looking from inside the system there is little reason to think that a Cause representation exists at this level.

Clauses also seem misplaced on this level. The presence of Clauses suggests the ability to compose Things and Differences. However, at such a low level, where it is not clear that learning occurs, it seems meaningless to talk about combinations of Things and Differences since it seems implausible that a system exists that would take advantage of such combinations.

Learned Reactions

Looking through the lens of Learned reactions, we find that a different notion of the Causes and Clauses framework can be applied. We see as context for this level the combination of unsupervised and supervised learning, as will be explained below. We will again talk about Things as perceptual data, but this time as residing on a medium-level. Such a Thing might well be thought of as a prototype or the principle components that represent “objects” whether they be physical or abstract. A Thing would be well described using a noun. Things could be thought of as composed of lower-level perceptual data into a more coherent and general symbol, such as the concept of a chair can be the combination of many views of chairs plus a notion of how a chair is used. As before, Differences are medium-level sensory descriptions of dissimilarity.

The preceding description of Things and Differences captures the unsupervised learning aspect of the Learning level. This is demonstrated upon consideration of the way that these representations would be acquired. The most reasonable way that this could work is through a self-organizing process that categorizes or clusters similar percepts. Thinking about Things as feature vectors and Differences as dissimilarity lends itself to self-organization because algorithms that accomplish self-organization use those data types.

A different kind of learning, supervised learning, suggests a different way of viewing Things and Differences in this level. But while unsupervised learning is content to occur without much reference to the temporal realm, supervised learning cannot work without it. In particular, we choose to add a temporal dimension, as the sequence of presenting a stimulus and observing the response is meaningless without it. Once we can stamp Things and Differences with times, we can begin to look at Things as the state of an object, and Differences as trajectories that evolve over time. Clauses in this level can be described as state-trajectory-state transitions.

This description covers some ground towards describing supervised learning, but leaves out a significant notion, that of *Reward.* We have discussed above how reward, in the form of some kind of natural selection, provides the pressure to decompose the world in increasingly powerful ways. Here, reward, in its most abstract sense is simply a scalar value assigned to an action (or strictly, a state-action pair) as a means of altering the likelihood that the system will repeat that action under similar circumstances. As we will try and show below though, the notion of reward takes on increasingly varied significance at higher levels.

In particular, for our explanation, it is clear that supervised learning is meaningless without some way of deciding between the things you want to remember and the things you don’t mind forgetting. We feel that the notion of reward fits that purpose, and thus we will carry it through the following sections alongside the Causes and Clauses framework and analyze how further levels can be thought to involve Reward.

At the end, we are left wondering if Causes can be integrated in at this level. A starting point for thinking about Causes would be to consider them as explanatory agents responsible for Differences, but this description falls short for a few reasons. For unsupervised learning, it does not make much sense to attribute an agent to the difference between objects. Asking why a prototypical apple is different from a prototypical orange isn’t likely to add much useful information. But in supervised learning, it also doesn’t make sense to ask why one state is different from another state. In any kind of supervised learning, the “agent” responsible for the state of the system being different is always the same; it is the sheer fact that a Difference between the two states exist. In the example of the back propagation algorithm, the weights of the synapses of the network will be caused to change simply because the network returned a value which didn’t match a target. In general, we don’t think of a feed forward network as representing *why* the synapses change; it is only important by how much they should change. We feel that this same philosophy generalizes to all of the supervised learning methods used at this level. Because of this, it still does not appear that a Cause can fit into Model Six at this level.

Deliberative Thinking

To understand the notion of thinking on the Deliberative level, let us begin by defining what sort of Clause we see resulting from this level, and work backwards to explain what parts are required to construct it. We view this level of thinking as planning, where “sub-plans” are chained together in a sequence for some purpose. In order to investigate further, let us consider first what these “sub-plans” might be, and then consider what we can say about the purpose that planning serves. We believe that a profitable way to think about “sub-plans” is as state-trajectory-state transitions, as we established were the Clauses from the Reflective level. Generally one thinks of a plan as: “first I do X1, then I do X2, finally I do X3”. We envision this as a sequence of three transitions, where a single transition involves going from a state where Xn isn’t done to a state where Xn has been done. The trajectory in this case could be described as “doing”. For our purposes, it makes sense to call these transitions the Things for this level. Additionally, we can imagine Differences as dissimilarities between transitions.

What about the purpose of a plan? It is now important to consider the role of Reward on this level. One way to think about how a plan is motivated is by the fulfillment of some kind of need. Thus, we can define Reward as the degree to which the current plan is determined to fulfill the present need. With this in mind, we can be more specific about the process of Deliberative thinking. The process can now be described as generating test chains of transitions, evaluating those test chains on the basis of Reward, discarding the parts that don't work and adding parts that do, resulting in a final plan.

To aid in the process of Deliberative thinking, it finally seems important to introduce the concept of Cause. Here the utility of a Cause obvious. In order to make determinations about how to order transitions correctly, transitions need to be able to be described in terms of causality. Without a notion of Cause in a Deliberative thinking phase, we could never tell why anything happened, and would be stuck trying to increase our Reward without any solid notion of how to go about doing it. We must know that one transition will make another one occur in order to plan effectively, and this is at the heart of what a Cause is.

Reflective Thinking

We now move onto the Reflective thinking level. This layer's purpose is to evaluate the plans created in the Deliberative level in light of the execution of the plan. Things in this level are most profitably associated with the sequences of transitions created as clauses from the Deliberative level. We find that Reward continues to be relevant at this level. Here, Reward can provide a record for how closely the plan matched up with its execution. We might imagine a special kind of difference engine, perhaps called a difference detector, which measures the difference between plan and execution, rather than trying to eliminate it. The closer that the result sticks to the plan, the higher the Reward gets. In this way, we perceive sticking to our plans as favorable, and deviating as negative and frustration.

What are the Differences? Because it is difficult to conceive of Differences as comparisons between plans, let us employ the image of a program which carries out a matching operation on plans. Its role is to compare the structures of different plans and to output the ways in which the two structures do not match, by process of elimination with all the things that do match. These leftovers can be considered to be the Differences in the system.

Do Causes continue to have a place in this formulation? In fact, they do. We can think of Causes as Things which caused plans to be executed in particular ways. If a plan failed to work for a few salient reasons, the Things that can be associated with those reasons should be considered Causes.

The resulting Clauses from this level are best described as scenarios. You had plan A, and got result B, and felt Reward C. This describes a miniature story that can be saved away in episodic memory, and retrieved as a unit of its own and compared as such.

Self-Reflective Thinking

Self-reflective thinking is a process that centers on a mental model of self. The model can be thought of as an experimental testbed where the imagined results of being in certain situations can be explored. This allows a person to attempt to achieve something of an "objective" perspective of themselves and what they are doing. On this level, a Thing is most wisely attributed to the different scenarios that you use to experiment with this model. Differences, understandably, are the ways that the scenarios are dissimilar, and could also be explained using a difference detector. Reward measures the extent to which you are 'satisfied' with the results of model, given the scenario you put into it. Causes are still relevant, and can be associated with the reasons you give for making a particular decision based on a particular stimulus. Finally, Clauses can be associated with a type of memory that captures a record of the experiment and its results.

Self-Conscious Emotions

Lastly, we examine the level of Self-Conscious emotions. Its role is particularly abstract, and can best be defined as 'what you think other people think (or would think) about your actions'. While the previous level involved a single model of mind, this level requires mind-models for all of the people whose thoughts you are considering. Here, it makes most sense to think about the opinions that those different mind-models would produce as the Things, and comparisons between those opinions as the Differences. Reward is still tightly bound into this notion, as one could imagine that more critical opinions would garner less Reward than praising opinions. The Causes for these opinions would be highly correlated with the way in which you have represented the minds of others. While you might attribute your own mistakes to circumstance, you might represent the minds of others such that their mistakes are attributed to their nature. Causes would likely be described as personality traits or past experiences each person you considered possessed. Finally, the Clause for this level seems to make most sense being assigned to the combined set of opinions that you receive back after querying your many mind-models.

Contributions

In this paper, we discussed the extent to which we could try and discuss intelligence in the abstract, and confidently make predictions about all intelligent systems. We argued for the decomposition of the world in the eye of a learner, as a means of compressing it and predicting it by 'splitting it at its joints'.

We took Minsky's Causes and Clauses framework and, treating it as almost orthogonal to the Model Six hierarchy, combined the two, to show how generative the combination could be. In particular, we discovered that the notion of Reward was integral to the confluence of these two ideas, and appears to serve both frameworks extremely well. By examining the kinds of representation that fit at each level, we proposed and redefined the original causes and clauses framework into a hierarchy of sorts that could be almost indefinitely extended to produce increasingly fine-grained ways of breaking up the world.