The Origin and Resolution of Ambiguities in Causal Arguments

Johan de Kleer
Xerox Palo Alto Research Center
Palo Alto, CA 94304

Abstract

The causal arguments that people typically use to explain the behavior of physical systems contain ambiguities and hidden assumptions which result from imposing a particular point of view on the behavior of the system. The causality of such an argument is an artifact of imposing this point of view. Usually there exist other equally “valid” but conflicting arguments based on the same evidence. The inherent local nature of causal arguments makes it impossible for them to capture the more global effects that are needed to resolve these ambiguities. However, their local nature makes causal arguments computationally simple to construct. This paper discusses these ideas in the context of electronics after first presenting a general theory of causal arguments. The causal rules that electrical engineers appear to use to reason about circuits are presented, and their use in constructing causal arguments for circuit behavior is discussed.

Introduction

This research attempts to articulate the nature of the causal arguments which are so prevalent in human explanations of physical phenomenon. In particular, this paper is concerned with exploring “mechanism-like” explanations for the behaviors of physical systems. Although mechanistic or causal arguments are ubiquitous, their structure and generation is rarely formalized. Both the knowledge and its associated calculus is usually tacit: there exists a shared body of common knowledge which is referred to when analyzing, explaining and understanding physical systems. Thus, this research can be viewed as articulating the tacit calculus used to understand and discover causal arguments.

This kind of research can have impact on education, cognitive science, and artificial intelligence. This tacit calculus is part of what we are really trying to communicate in the classroom. If we understood it better this education process could be made more efficient. Furthermore, a formal model of the informal tacit calculus provides the foundations for computer-based instructional systems [Brown, Collins & Harris 78]. From a cognitive science perspective, causal reasoning is an instance of a more general type of reasoning about physical systems, namely envisioning [de Kleer 77]. The study of causal reasoning provides another set of distinctions that helps clarify envisioning. This common sense tacit calculus is exactly what most modern problem solving systems have been lacking, and thus a better understanding of the tacit calculus can lead to more robust knowledge based systems.

In NEWTON [de Kleer 77] an initial theory of envisioning was developed which focused on a kind of qualitative simulation of the physical system of the roller-coaster. An example of a problem NEWTON could deal with was:

Figure 1: Will the cart reach X?

NEWTON approaches this problem by simulating the behavior of the cart utilizing only a few qualitative features such as the concavity of the surfaces and the direction of the velocity of the cart: “The cart rolls down the track and starts rolling up the inside of the loop. At this point it may roll back and oscillate. As it approaches the top of the loop it may fall off...” Since the specific values of the initial velocity and heights of the roller-coaster are not taken into account, the qualitative simulation cannot resolve the two critical ambiguities: (1) will the cart oscillate, (2) will the cart fall off. In order to resolve the ambiguities NEWTON analyzes their underlying assumptions and uses this information to reformulate the qualitative problem as quantitative one for subsequent quantitative analysis. For example, if the cart reaches X and does not oscillate, it must be assumed that the velocity did not go to zero. This assumption then guides the quantitative analysis to look at the velocity of the cart on the initial segment of the loop. After these ambiguities have been resolved, the behavior of the loop-the-loop system of Figure 1 can be correctly simulated. The original qualitative simulation of the envisioning could not resolve the ambiguities and thus had to consider all the alternatives, which quantitative analysis had to later resolve. Therefore, envisioning can be viewed as the construction of a space of possible simulations which some other knowledge has to distinguish among.

A goal for this research is to develop a more sophisticated theory of envisioning. One of the serious problems that arises in extending the theory is that the behavior of more complex physical systems is typically
Electronic circuits provide a rich domain to explore these issues since they are excellent examples of constraint systems (all of the network laws are expressed as constraints). Yet humans have extensive experience in dealing with them. Therefore after the general theory of causal reasoning is presented in the next sections a causal reasoner for electronic circuits will be developed.

The use of this common sense tacit calculus leads to more powerful artificial intelligence programs. Causal explanations describe how the behaviors of the individual constituents contribute to the overall behavior of the system. This knowledge is important for understanding, designing and troubleshooting designed systems. For example, in the case of electronic circuits a complete algebraic analysis of even simple circuits can be computationally prohibitive, but knowledge of how the individual components contribute to the circuit's composite behavior can significantly improve the efficiency of the analysis [de Kleer & Sussman 78]. For example, an integrated circuit operational amplifier contains a large number of transistors, but few of them are situated on the main signal path. For many calculations, the effect of these auxiliary transistors on the signal can be ignored or accounted for by much simpler transistor models. The use of these simpler models significantly reduces the complexity of the algebraic analysis.

The causal explanation identifies which transistors are crucial to the behavior and which are not. Causal reasoning also plays a fundamental role in identifying the faults responsible for symptomatic behavior and in localizing faults at a shallower level of detail before entering the more expensive deep analysis [Brown 76] [de Kleer 76]. Early designs can be checked to see whether they have any hope of achieving their desired behavior, and the sections which are critical to the desired behavior can be identified for special attention [McDermott 76].

This research differs from related work by Freiling [77] and Rieger & Grinberg [77] by focusing on the distinction between the physical system that manifests the behavior and the abstract mechanism by which the system achieves that behavior. Rieger's theory has no representation of the system that his cause-effect diagram is a description of. In my approach causal reasoning is, in effect, a method to construct a description of the abstract mechanism from the description of the physical system. This approach has a methodological advantage over Rieger's since it eliminates much of the arbitrariness from the representation of any particular mechanism. The current research has not progressed to the point that it is capable of producing CSA-like description of the mechanism; this is a logical next step. In [de Kleer 79] the theory of causal reasoning outlined below is used as the basis for a recognizer, QUAL, which takes a description a circuit and produces a description of the mechanism by which the circuit achieves its behavior.

Theory

The general form of a "mechanistic" argument is a sequence of events occurring in the functioning of the physical system where each event can be causally related to events earlier in the sequence. Each event is an assertion about some behavioral parameter of some constituent of the system (e.g., velocity at a point or current through a terminal). The sequential argument always reads as if it is temporally ordered.

![Figure 2: The Schmitt Trigger](image)

"... An increase in \( v_1 \) augments the forward bias on the emitter junction of the first transistor, thereby causing an incremental increase in the collector current, \( i_C \), of that transistor. Consequently both the collector-to-ground voltage \( v_1 \) of the first transistor, and the base-to-ground voltage of the second transistor \( v_2 \), decrease. The second transistor operates as an emitter follower which has an additional load resistor on the collector. Therefore, there is an increase in the emitter-to-ground voltage \( v_2 \). This decrease in \( v_2 \) causes the forward bias at the emitter of the first transistor to increase even more than would occur as a consequence of the initial increase in \( v_1 \) alone...." [Harris et al. 66, p.68]

A causal argument consists of a sequence of assertions about system constituents each of which hold as the consequence of previous assertions. In a causal argument, \( A \) is a consequence of \( B \) means \( A \) is caused by \( B \). The causal argument "... An increase in \( v_1 \) augments the forward bias on
the emitter junction of the first transistor, thereby causing an incremental increase in the collector current, ..., is a sequence of two assertions: \( V_I \) increases, \( I_C \) increases. These are the two events of this causal argument. The deduction of one event from another is determined by the causal rules of the device models. In the above example the model for the first transistor is one in which increased emitter potential causes increased collector current.

The causal argument makes reference to an underlying device topology which describes the system's constituents and all possible interactions between them. Each event of a causal argument is an assertion concerning some constituent, and its antecedents refer to other events about constituents which are adjacent in the device topology. This is a consequence of the local nature of the causal models. A particular causal argument selects only a subset of the device topology as the information bearing connections. This subset is called the causal topology.

The local nature of the rules is the source of much of the power of causal arguments, as well as some of the problems. It ensures that every constituent (e.g. transistor, resistor, etc.) of a given type can be modeled the same way without regard to its surrounding constituents. Therefore the number of causal models is very small. This ensures that at least one and often many causal topologies can be constructed for any system constructed out of known constituents.

The potential ambiguities in a causal argument stem from (1) the qualitative nature of the rules, (2) the local nature of the rules, and (3) the inability to distinguish between the causes and effects among the events. Although the Schmitt trigger is inherently a constraint system and roller-coaster appears to be inherently sequential, the forms of the mechanistic arguments which explain their behavior are very similar. This was achieved by the engineer imposing a mythical causality on the behavior of the system which then admitted a temporal sequential argument.

The explanation for the Schmitt trigger made a number of unsubstantiated assumptions aside from the choice of transistor models. Why does the \( V_I \) increment appear across \( Q_1 \) instead of \( R_E \)? Why does the voltage \( V_I \) drop since \( Q_2 \)'s turning off should raise it? Why is the current contributed by \( Q_2 \)'s turning off more than the current taken by \( Q_1 \)'s turning on? There are many values for the parameters for which the circuit cannot function at all. The arguments are only rationalizations of the observed behavior (observed by actual measurements or stated in the textbook). This does not detract from the usefulness of the explanations: no explanation ever accounts for every detail of the behavior. The usefulness of an explanation does not depend on how complete or correct it is, but whether the explanation is sufficient for the purposes it is applied to. One of the aims of this research is a taxonomy of different types of rationalization and when they are useful. With such a taxonomy, for example, we can hope to make some progress in the area of automatic summarization and explanation of how complex systems work.

All of these ambiguities can be traced to assumptions made by the local rules. In order to see this, the discovery, or generation, of causal arguments must be considered in more detail.

Generating Causal Arguments

The apparent temporal order of the events of a causal argument allows it to be viewed as the description of a simulation of the system's behavior. This simulation can be easily repeated by applying the causal models to the underlying causal topology of the argument. The method for originally discovering a causal argument is also a simulation but without the knowledge of this causal topology. This generation simulation splits at every point where the device topology leads to different causal topologies thereby producing a collection of different causal arguments each with its own distinct set of underlying assumptions. The same causal models are used in the resulting causal argument as in the generation process. This is a stronger locality claim than made in the previous section since it states that the models used for the discovery of the original argument can only refer to information local in the device topology. This is surprising since one would expect that the generation simulation would require more detailed models.

This simulation method of generating causal arguments is computationally quite simple. The finite device topology ensures that the simulation must terminate, and the local nature of the device models makes the simulation of devices simple, as well as ensuring that every system constructed from the modeled devices can be analyzed. (The generation simulation can be compared with conventional forward deduction, except that the deductions are severely limited by the device topology and that there is no negation.)

In order to construct a causal reasoner for a given domain three issues have to be addressed: (1) What is the device topology? (2) What are the rules of the device models? (3) What are the possible assumptions, and how should they be dealt with?

The remainder of this paper explores these issues in the context of electronics.

In electronics, unlike most physical systems, the device topology can be determined directly from a circuit schematic. The choice of electronics as a domain to explore causal reasoning simplifies this first issue considerably. As the models are presented a number of different assumption types will become evident, but all the types of assumptions are dealt with in the same way.

Device Models for Electronics

The classical engineering models that are used to model the behavior of electrical devices are widely agreed upon. However, the causal qualitative models that people use to reason about circuits are not. In fact, these qualitative models
are rarely articulated, even though the tacit models that underlie people's arguments appear to be very similar. The following discussion glosses over the difficulties of identifying these heretofore tacit models. For a more detailed discussion of how these models were arrived at and alternative models see [de Kleer 79].

The causal explanation of how a circuit works is a qualitative description of the equilibrating process that ensues when signals are applied to the circuit. The behavior of the Schmitt trigger was described in this way. This will be called incremental qualitative (IQ) analysis. Since most circuits are designed to deal with changing input signals, it is not surprising that the main purpose of most circuits is achieved incrementally. For example, an amplifier must amplify changes in its input, digital circuits must switch their internal states as applied signals change, and power-supplies must provide constant current or voltage in the face of changing loads and power sources.

Incremental qualitative arguments rarely need to refer to more than the sign of the derivative which indicates whether the signal is increasing or decreasing. This requires an algebra of four values: "t" signal is increasing, "0" signal is not changing, "i" signal is decreasing, and "?" signal is unknown. The arithmetic of this algebra is very simple:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>t</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>t</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>0</td>
<td>0</td>
<td>t</td>
</tr>
<tr>
<td>?</td>
<td>t</td>
<td>t</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 1 : x + y

Only addition and subtraction are important, and no other operations are ever used.

The approach for constructing the models is to start with the classical constraint models, and reformulate them preserving only the sign of the derivatives of the variables. Ohm's law has a particularly simple formulation:

\[ V = iR \]

Figure 3: Ohm's Law for Resistors

\( \Delta x \) refers to the sign of the derivative of \( x \). Currents are defined to flow into devices away from nodes. Kirchhoff's Current Law (KCL) applies to components so that the current into #1 is equal and opposite the current through terminal #2.

The IQ model for Ohm's law is:

\[ \Delta v \leftrightarrow \Delta i \]

The rule specifies that the derivative of the current must be of the same sign as the derivative of the voltage. Since the resistor has no preferred causal flow direction this rule must be bilateral. This action is specified by the \( \leftrightarrow \) operator.

The behavior of nonlinear devices can be modeled by a small number of linear regions, or states. Since the correct state cannot be determined when a nonlinear device is first examined by the simulation, the use of a causal rule for any state can only be made under the assumption that the device is operating within that particular region.

We now have enough rules to analyze some simple circuits:

The following causal argument is generated for this circuit (assuming the transistor is on):

Input voltage goes up (Given)
Collector current increases (Transistor rule)
Current through the resistor increases (KCL)
Voltage across the resistor increases (Resistor rule)
Voltage at output decreases (KVL)

In order to generate the causal argument the conventional
Kirchhoff's Laws were utilized. Kirchhoff's current law (KCL) specifies that the current flowing into any node or device is zero. Since there are only two devices connected to the output node, KCL states that the current flowing out of the collector equals the current flowing into the resistor. Kirchhoff's voltage law (KVL) specifies that the voltage around any loop is zero. Since the supply voltage is unchanging, KVL states that the voltage change across the resistor must be equal and opposite the voltage across the transistor.

Heuristic Rules

The rules discussed so far are sufficient to deal with many circuits. The only assumptions that are made involve state choices and therefore every causal argument that makes the correct state choices is guaranteed to be valid. Unfortunately, most circuits cannot be completely analyzed. One such case occurs when a transistor's collector is connected to a number of circuit fragments:

Figure 6: An Unanalyzable Circuit Fragment

An increased base voltage causes an increased collector current. Since there are a number of devices connected to the collector node, the current through these devices cannot be determined. Furthermore, the transistor model provides no information about the voltage at the collector node. Thus the simulation halts with the voltage and currents at the node left unassigned. In quantitative analysis this would be a point at which to introduce a variable [Stallman & Sussman 77].

This situation arises quite commonly in engineers' analyses. The method for dealing with it can be summarized in one heuristic: the (IQ) value of the potential at the node (the voltage with respect to ground) is opposite to the (IQ) value of the current drawn from the node. An instance of this heuristic is: "The increasing current pulls down the node." In the above circuit fragment, the increased collector current causes the voltage at the collector node to drop. This heuristic rule is makes the assumption that the circuit around the node is behaving as a positive resistance. This assumption can be violated.

A second unanalyzable circuit fragment occurs at the input of the Schmitt trigger:

Figure 7: Schmitt Trigger Input

The input signal is applied between the base of the transistor and ground (i.e., across both the transistor and the resistor). Since this voltage does not appear across any component in isolation, no causal argument is discovered for the behavior of this circuit. A second heuristic is used to deal with this case: whenever a voltage is discovered between a device terminal and ground, this value is applied directly to the device. This heuristic assumes that the first signal discovered at the input of some device dominates the signals at the other terminals.

The first heuristic is called the KCL-heuristic, and the second the KVL-heuristic. These two heuristics combined with the device rules are sufficient to explain the behavior of most circuits. The central remaining difficulty is that this causal calculus does not necessarily ascribe a unique behavior to the circuit being analyzed. These ambiguities result from assumptions made by the heuristic rules. The assumptions come from three different sources:

1. The nonlinear nature of the devices.
2. Transitions between the linear states.
3. KCL- and KVL-heuristics.

Ambiguities and Assumptions

A second phase of the analysis considers the different possible causal arguments (ambiguities in the behavior) that a circuit can have in order to determine which assumptions are relevant. In order to do this each event of a causal argument is tagged with the assumptions that were made in the simulation to get to it. This tagging is illustrated by the following more formal description of the causal argument for the Schmitt trigger. Each voltage or current is described by a short descriptor, a value, a list of assumptions, and a justification. The assumption list is enclosed within "(...)". An assumption concerning a device's state is described as (device <state>), a KCL-heuristic assumption is described as (terminal <device>), and a KVL-heuristic assumption is described as (current <device> <terminal>).

(VOLTAGE INPUT GROUND) = \downarrow
Premise.

(VOLTAGE B1 E1) = \downarrow (Q1 B)
KCL-heuristic [Q1 B]

(CURRENT C Q1) = \downarrow (Q1 ON) [Q1 B]
\partial V = \partial IC for Q1

(VOLTAGE C1 GROUND) = \uparrow (C Q1) [Q1 ON] [Q1 B]
KCL-heuristic [C Q1]

201
These assumptions play a critical role in two rather complex kinds of qualitative reasoning, which are presented next.

Points of View

This section and the next contain a very brief discussion of the roles of interpretations and teleology. Both involve surprising subtle and complex issues which are difficult to present in the space available. See [de Kleer 79] for a detailed presentation.

A collection of assumptions can define a point of view on the behavior of the circuit. Each different collection of assumptions suggests a different way in which the circuit can work by selecting a different set of events which depend on the assumptions. There a number of conditions that a collection of assumptions must satisfy before it can be reasonably considered a point of view. For this reason the notion of interpretation is introduced. An interpretation of a circuit’s behavior is a collection of assumptions which:

1. is noncontradictory -- does not select contradictory events.
2. is maximal -- every event that can be added must be added.
3. justifies heuristics -- all the assumed positive resistances must behave as such.

These three conditions rule out most collections of assumptions, and the remaining interpretations are very few (never more than twice the number of transistors in the circuit). Often there is only one interpretation.

The Role of Teleology

Every interpretation produced by this procedure describes a behavior that could be valid for some assignment of circuit parameters (i.e. specific resistances, gains, power supply voltages, etc.). The interpretations can usually be disambiguated by appealing to the teleology of the circuit. Since circuits are designed artifacts -- systems whose behavior is to achieve some particular purpose -- they have a teleology. By knowing the purpose of the device the interpretation whose behavior is consistent with this purpose can be selected as the correct one. For certain classes of circuits, just the knowledge that the circuit has some purpose can be sufficient to identify the correct interpretation.

In the case of the Schmitt trigger the analysis discovers four interpretations:

1. correct.
2. approximately interpretation [1] but without feedback.
3. signal reversing the feedback path.
4. approximately interpretation [2] and [3].

(Interpretations [2] and [3] do not violate the maximality condition for interpretations, but in order to see this a detailed examination of the assumptions involved is required.)

Conclusions

The power of the causal reasoning (envisioning) comes from the fact that it is complete, limiting and articulate. Envisioning is complete for the class of circuits QUAL considers since it is capable of simulating every possible behavior. Therefore, any behavior which the envisionment does not predict as a possibility cannot happen. Envisioning is limiting in that it generates only a small number of ambiguities. Finally, envisioning articulates the source of the ambiguities so that other knowledge can be used to deal with them. In the case of electronics this other knowledge is teleological. Without any one of these properties the calculus would be useless. For example, without the completeness property no necessary relation exists between the envisionment and what is actually the case. If the envisionment does not articulate the source of the ambiguities, other knowledge cannot be used to resolve them.

NEWTON [de Kleer 77], a program to solve physics problems in the roller coaster domain, is also based on envisioning. The envisionment for a roller-coaster problem consists of a sequence of qualitatively described scenes indicating how the roller-coaster moves along the surface. NEWTON invokes mathematical laws to quantitatively resolve
the ambiguities, while QUAL resolves the ambiguities quite differently by appealing to the teleology of the system. Both programs are based on the same theory of envisioning, but resolve the ambiguities in radically different ways. Although ambiguities appear at first to be an undesirable side-effect of a theory of envisioning, the examples from physics and electronics illustrate that they play an important role in understanding complex systems. The assumptions underlying the ambiguities provide the key to reformulating the analysis so that other knowledge can be profitably employed.

In the case of the roller coaster world the causal rules are quite obvious and are determined by the behavior of the roller-coaster cart through time. However, in the case of electrical systems it is very difficult to determine a time order on the events, and the causality between events is largely imposed by the understander. In this case the causal rules can only be determined by studying arguments that people actually use to understand circuit behavior.

Acknowledgements

I would like to acknowledge the help of my thesis advisor G.I. Sussman. I had many enlightening discussions with J.S. Brown concerning the content of this paper. A forthcoming joint paper with J.S. Brown deals with the cognitive aspects of this work.

This paper describes research done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's research is provided in part by the National Science Foundation under Grant MCS77-04828.

References

[Brown 76]

[Brown, Collins & Harris 78]

[de Kleer 76]

[de Kleer 77]

[de Kleer 79]