Perceiving Complex Causation Through Interaction

Colin Ware
CCOM, University of New Hampshire
Durham, NH. USA

Abstract

When we interact with a touch screen computational device we have the strong subjective impression that we are directly causing changes that occur on the screen. For example, sliding a finger on a screen causes scrolling of the information. But the current method for showing causal relationships derived from models is to use a causal network diagram with nodes representing entities and arrows represent causal relationships between those entities. Sometimes arrows are labeled to weight the connections. In such diagrams there is no immediate visual impression of *causal* links, just the perception of connections and the arrowhead symbol. Interactive touch screens would seem to offer the potential for creating interactive diagrams where the causal relationships are provided in a perceptually immediate and unequivocal fashion. This paper explores methods for creating interactive diagrams using multiple touches that go beyond simple positive causation to express complexities such as causal effect enhancement, causal effect reduction and causal effect blocking. A design rationale is presented with special attention to temporal constraints. Results from an evaluation study suggest that the design can be understood with minimal instruction by most people.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—Human information processing.

Additional Key Words and Phrases: Causality, Animation.

1. Introduction

If we pull on a string and immediately something moves, we infer a causal relation between our action and the result. Toss a stick in a bush; if an animal runs out we assume we caused that too. The acts of grabbing, hitting, pushing and squeezing all result in direct contingent visual changes in the state of the world. It is hypothesized that it is through such temporal contingencies that infants gain a basic understanding of the state of the world (Cohen, 1998). Furthermore, cognitive scientists propose that such experiences form the substrate on which even very abstract concepts are built (Barselou, 1999; Pesher et al., 2004).

In prior research, only a very limited vocabulary of visual causal effects has been systematically explored, mostly derived from Michotte's (1963) extensive studies of positive causative effects. The goal of the research reported here has been to expand the available vocabulary of visually compelling and easy to learn interactive representations of causal effects, to include negative causation, amplification of causal effects, and blocking.

We begin by exploring some of the underlying perpetual and cognitive theories that motivate the approach taken, followed by a discussion of related prior research in the field of data visualization, before moving on to present a design rationale based on the availability of multitouch screens, temporal constraints and the need for a degree of abstraction.

1.1 Theory

The theory that cognitive concepts are based on sensory experiences has a long history, being set out by John Locke (1690) in the 17th century and even earlier by Aristotle. This theory fell out of favor in the 1980s and 90s but has, relatively recently, undergone a major renaissance. For example, Barselou (1999) argues that sensory experiences of time varying events are stored as neural activation sequences, and that these sequences act both as memories and as executable processes that can be used in future activities. In addition, these processes become the substrate of abstract reasoning about events in the world (also see Glenberg, 1997). Providing supporting evidence, linguists such as Pinker (2007) and Lakoff and Johnson (1980) point to the enormous richness of spatial and temporal metaphors in thought, as revealed by language, showing that even highly abstract concepts are often based on perceptions that have a basis in the spatial and temporal physics of everyday life.

A distinction is sometimes made between *amodal* and *modal* theories of causality (Wolff, 2007). Amodal theories propose that the brain encodes concepts such as causal relations in ways that are independent of sensory experience. An example is the *Baysian inference* theory, based on the work of Pearl, Tenenbaum & Griffiths (2001) and others. This holds that causality is inferred from statistical contingencies. If one event consistently follows another, it is argued, causality may be inferred. The objection that temporal contingency is not always due to causality has been addressed by statistical methods involving a third variable. The major problem with this as a psychological theory is that statistical inference requires large numbers of events to have been observed, but in fact causality is often inferred from a single observation.

A second example of an amodal approach is the *counterfactual* theory. This holds that we believe something to be caused by something else if, by removing the causal factor, the causal chain would be broken. However, as Sloman (2005) points out, the distinction between *cause* and *enable* is not specified by counterfactual reasoning. For example, the presence of an unlocked bicycle might enable an individual to escape an angry mob. We would generally hold that the angry mob, not the bicycle, caused him to flee. We would not say that the angry mob enabled him to flee, nor would we say that the unlocked bicycle caused him to escape. Yet without the unlocked bicycle escape would not have occurred.

As an alternative to amodal theory, *sensory experience* theories hold that causal concepts are based on the neural processing

---

*email:cware@ccom.unh.edu

CAe 2013, July 19 – 21, 2013, Anaheim, California.

sequences that occur when we experience and interact with the world. They are tied to the sensory modality of the formative experiences, and it is claimed that even abstract causal concepts are based on a kind of approximate modeling derived from everyday physics. Wolff (2007) calls this the *physicalist* theory. Leslie (1994) suggested that concepts relating to physical causation are processed by a primitive *theory of bodies* that schematizes objects as bearers, transmitters and recipients of primitive encodings of forces.

A basic assumption of physicalist theories is that physical causation is cognitively more basic than nonphysical causation, such as social or psychological causal factors. Supporting this is evidence that our ability to perceive physical causation first develops in infants at around 3 to 4 months, earlier than the ability to perceive social causation which occurs around 6 to 8 months (Cohen, et al. 1998). In addition, Wolff (2007) showed that a dynamics model is accepted as a representation of social causation suggesting that visual representations of causal links should be based on simple physics and not, for example, on little animated characters conveying causal information.

One of the consequences of the physicalist theory is that causal effects can be perceived as direct and immediate. Michotte's (1963) extensive experiments with moving lights showed that under timing and path constraints moving one moving patch of light could be made to phenomenologically *cause* a second patch of light to move. Indeed we perceive such casual effects every time we see an animated cartoon.

An experiment by Wolff (2007) lends support to the physicalist theory being applicable to combinations of causal factors. Wolff set up simple animated sequences using a boat on water containing a person facing forward and a representation of a bank of wind fans. The boat began travelling in a straight line then changed its course at the same time that the fans were activated in the animation. On the new course it rammed into a cone placed in the water. Two forces were involved in the model, the propulsion force of the boat engine and the force of the wind created by the fans. In some of the conditions the change in direction was consistent with the sum of the two forces; in others it was not. The evaluation was by means of a fill in the blanks, sentence completion task. The results showed that participants were much more likely to choose words saying that the fans *caused* the boat to hit the cone if the change in direction was consistent with a combination of the two forces. In cases where wind force could not have caused the animated change in direction, subjects did not attribute a causal effect to the fans.

The Wolff study shows that people can reason about the interaction of physical forces, but it does so using concrete animated objects. The purpose of the work presented here is to find more abstract ways of representing combinations of causal effects in interactive diagrams that can be rapidly and unambiguously interpreted. A useful method should be readily understood with minimal instruction, yet be sufficiently abstract that it can convey casual interactions in a range of different domains, such as medicine and finance.

The degree of abstraction is a critical variable in casual diagram design. An overly concrete representation will interfere with domain knowledge and may impede learning. The work of Goldstone (2005) showed that having subjects interact with simple physical models can promote learning of sophisticated concepts. However, it also showed that an appropriate level of graphical abstraction is needed. In a study of a method for teaching the concept of simulated annealing, he used a graphical representation with points bouncing off surfaces and eventually falling into a valley that represented the optimal solution. When realistic bouncing basketballs were used instead of bouncing points the transfer of knowledge to another problem was reduced.

1.2 Prior Work on Causal Graphs

Ware et al., (2000) defined a visual causal vector (VCV) to be a graphical device conveying a causal influence from one node in a causal network diagram to another. They explored three different styles: 1) a ball emitted from one node travelled to another, causing it to oscillate, 2) a rod, like a billiard cue, extended from one node, striking another and causing it to oscillate, and 3) a traveling wave emitted from one node, travelled to a bar shaped node, causing it to rise up.

The VCV concept has been extended. Bartram and Yao (2008) showed that causal strength can be efficiently encoded using the amplitude of node animation and they also shows that causal sequences could be reliably expressed. Kadaba et al. (2009) demonstrated a variety of effects, including flow though causality where a causal effect was mediated by an intermediate variable and two (non-interacting) causal influences on a single variable. They adopted the moving ball VCV, with a ball conveying a causal effect from one node to another, which they called an animated 'bullet'. High and low magnitude effects were portrayed by varying the amount of size change of a node that was the target of a bullet vector. Positive and negative causation was shown using + and - symbols and by having the target node either increase or decrease in size. Because of the use of the +/- symbols the results cannot be taken as clear evidence for a pure physicalist interpretation but the results from their evaluation study suggested that the interactive diagrams were easy to understand.

Neufeld et al. (2005) developed an interactive method for visualizing statistical causal relationships. In this case the goal was not so much to get users to perceive cause and effect directly in the Michottian sense, but rather that they should understand how the statistical probabilities changed depending on the particular values of causal factors. An arrow diagram was used to show the causal relationships between nodes. The interaction was through slider bars attached to the nodes, enabling the user to move a mouse to adjust the value on a node. Contingent effects could be observed as other slider values moving up and down in response. However, in the absence of a user evaluation it is not possible to know how accurately the contingent changes were interpreted.

1.3 Research Goals

The purpose of the research described here has been to explore the design possibilities of using multitouch interactions combined with moderately abstract representations for conveying causal interactions having the following form: A has a causal effect on B, C modifies that effect. For example, being stressed is thought to increase the adverse effects of a flu virus. In this cast the flu virus (A) causes the infection (B) and stress (C) exacerbates the A → B effect. In other instances variable C might reduce the A→B effect; for example, if C represented flu medicine. In yet other cases the variable C might completely block the primary causal effect. There are many examples in biology of chemical agents blocking the uptake of other chemical agents. The notation C~(A→B) is used to generically denote instances where a secondary variable modulates a primary causal effect.

Throughout this paper A will represent the primary causal agent that has an effect on B, and C represents a secondary agent that modifies the first causal effect. Specifically, the goal was to represent six different causal interactions where a second causal agent modifies the effect of a first causal agent. These are listed below and given in the form of a diagram in Figure 1.

1. A has a positive effect on B and C enhances that effect.
2. A has a positive effect on B and C reduces that effect.
3. A has a positive effect on B and C blocks that effect.
4. A has a negative effect on B and C enhances that effect.
5. A has a negative effect on B and C reduces that effect.
6. A has a negative effect on B and C blocks that effect.

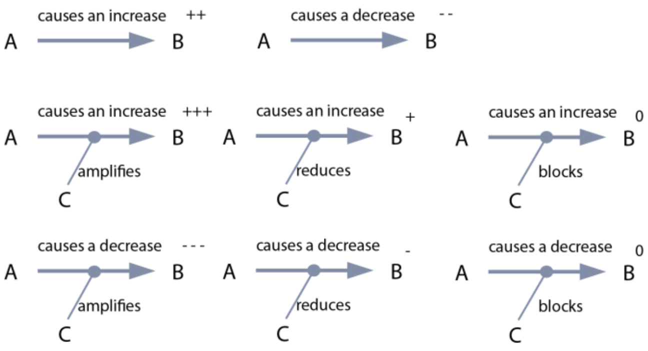

Figure 1. *The problem: represent six different causal interactions having the form A causes some B, allowing for both positive and negative causal effects, and C modifies the A→B effect, either enhancing, reducing or blocking it.*

2.0 Design Constraints

The design goal is to create a set of interactive representations building on the theory that causal concepts often (and perhaps always) derive from a set of naïve basic physical concepts originating very early in life as an infant first interacts with the world. A good solution to the causal representation problem must meet a number of design requirements. It must be compact so that it can be used in moderately complex diagrams with perhaps twenty or thirty nodes. It must have a reasonable level of abstraction. The causal interaction between variables must be discoverable by someone casually exploring the display. In the following sections a solution is developed together with its design rationale.

2.1 Level of Abstraction

In general, the greater the degree of abstraction of a diagramming convention, the more easily it can be used for a large variety of concepts. This is why arrows in directed graphs are so ubiquitous. They can stand for any of a huge range of directed relations. But the cost is that simple arrows convey very little information. They only tell us that some directed relationship exists between two entities; it might be causal or it might relate to the flow of information, physical forces or many other things. The advantage of a representation that conveys causality in a perceptually unambiguous way is that no effort is needed to learn this particular aspect of the meaning. If causality is perceived then cognitive resources are freed for other activities such as deducing the further implications of a causal effect. This can be especially useful in interactive museum exhibits and other learning materials where participants will not take the time to learn an unfamiliar notation. Nevertheless, good design will be as abstract as possible while maintaining the power of physical causality in the representation. The use of explicit animated agents, as in the work of Wolff et al (boats and fans) will be confusing in a display designed to teach people about biological pathways.

The design that was developed is as follows.

2.1.1 A moving streak of color as a visual causal vector

The VCV design chosen was neither the wave nor the travelling ball [Ware et al.,1999] as these were judged to be less abstract than was desirable. Instead a streak of color traveling in a wave along a spline curve was developed. This takes 0.33 seconds to traverse the length of the spline. While the streak of color does suggest something physical passing between A and B, it does not explicitly represent either a fluid wave or a physical particle, but can be given either interpretation. The traveling streak can be blocked, expanded or reduced to express a modifying factor.

2.1.2 Size changes for causal effect

Visual size is a ubiquitous and very general metaphor for quantity so size changes were used to express positive and negative causal effects. Positive effects are portrayed by means of size increases of the target node sphere. Negative effects are portrayed by means of size decreases of the target node sphere. However, a single instantaneous size change cannot work for a number of reasons. If someone repeatedly selects a causal factor node, the effected node will gain more and more size -for a positive effect- until it dominates the display space, or it will shrink to invisibility in the case of a negative effect. Also, there is a need for some degree of effect persistence to show modifying effects (C). There is a need both for an effect to somehow persist and for the diagram to return to a base state. These issues are discussed in more detail in the timing section.

2.2 Blocking and amplifying elements modifying the VCV

VCV links can have modifiers (see Figure 2). These are graphical elements designed to help convey causal enhancement, causal effect reduction and causal effect blocking of the A → B effect. The enhancement and reduction symbol consists of a rectangular node with a channel through it, straddling the primary VCV. This channel can either open or close based on a signal arriving from C. Simultaneously with this opening or closing, the primary VCV fans out to show enhancement, or narrows showing reduction (Figure 2a). The signal from node C to the blocking or amplifying element is also conveyed using a moving color streak VCV.

In the case of blocking, a disc appears when signal C reaches the junction point, and at the same time the VCV signal terminates at that point, visually failing to reach B. Examples of this design are shown in Figure 2.

In order for a user to perceive an interaction, such as amplification, it is necessary for the following conditions to be met: a) the user perceives some sense of a baseline effect representing the state of B without a causal effect, and b) the user understands the sign (positive or negative) and magnitude of the effect (A→B). It is only once these variables are perceptually

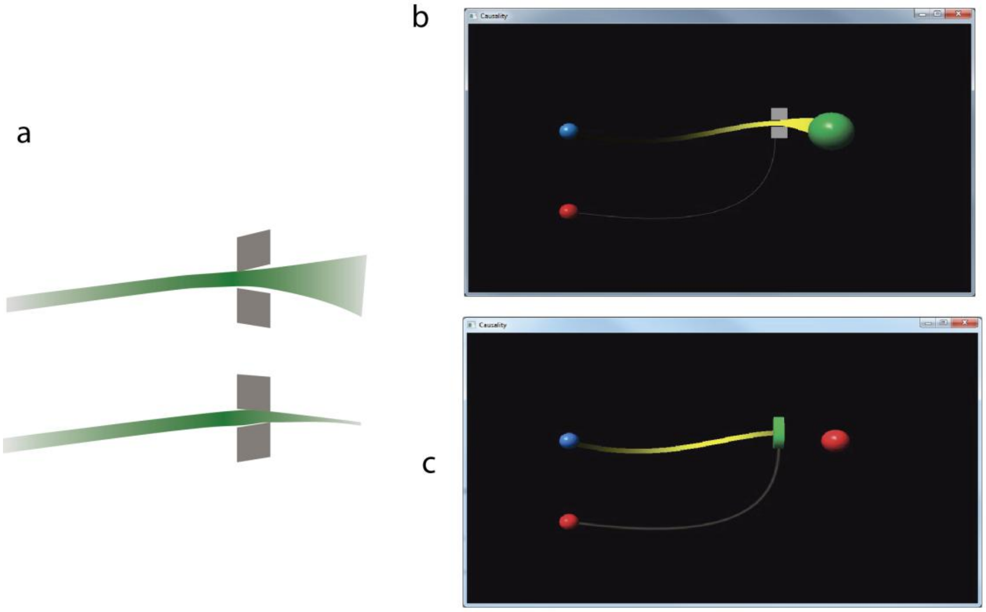

Figure 2. *(a) The graphical changes in the VCV to illustrate causal effect enhancement and causal effect reduction. (b) A screen showing causal effect enhancement. (c) A screen showing causal blocking.*

established that the effect of a modifying variable (C) can be understood.

This has temporal implications: A→B has to have already occurred before C~(A→B). Furthermore, in order that a secondary modifying effect be perceived, it is necessary that the modifying VCV arrive at a time when the primary effect is still somehow ongoing. In other words, A→B cannot be shown as a simple instantaneous pulse; it must persist so that the effects of the modifying variable can be observed, as illustrated in Figure 3.

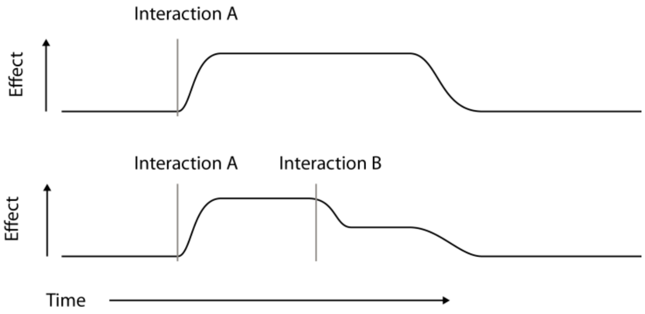

Figure 3. *For the modifying effect of a second interaction to become apparent, it is necessary that the first effect be visually persistent.*

The method for providing persistence was as follows. When the primary node (A) is touched a stream of yellow light streaks are emitted, each of which results in an increase in the size of the target node. Node effect decay is incorporated to take care of the problem of node B becoming too large.

Putting all of this together, the solution that was implemented is as follows.

1) Touching a node (A) representing a causal factor causes it to become highlighted with an attached ring of light, and the attached VCV conveys a series of 10 colored pulses at 3 Hz lasting for 3.33 seconds. Each light pulse travels along the spline curve of the VCV to the destination node (B).

On the arrival of a light streak, the recipient node has an immediate increase in size of 36 %. It also continuously decays to its mean size. It is useful to think of this in terms on an energy model described by the function

$$e_{t2} = (e_{t1}-b)0.3^{(t2-t1)} + b$$

where b is the baseline energy level (no inputs) and et1 is the energy at time t1. Assuming that there have been no additional energy inputs to the node in the interval [t1,t2] then the node energy at time t2 will be et2. Times are in seconds. The overall effect of the pulse series is to increase the node energy level and its size to a bit more than double. The overall visual impression is of a series of pulses that seem to pump up a rapidly leaking balloon.

In the case of a negative cause effect, the recipient node has an immediate decrease in size of (30%). It uses the same decay function and as a result shrinks to about 10% of its original size at the end of the pulse series.

The secondary causal factor (C) causes a modification to the A→B effect by acting on the amount of energy that is transmitted along the VCV. The effect of the modifier VCV is to increase or decrease the link signal (by about +175% or - 50 %). Only a single VCV pulse occurs as a result of touching C but this can be repeated. The effect of the VCV pulse on the amplifying element

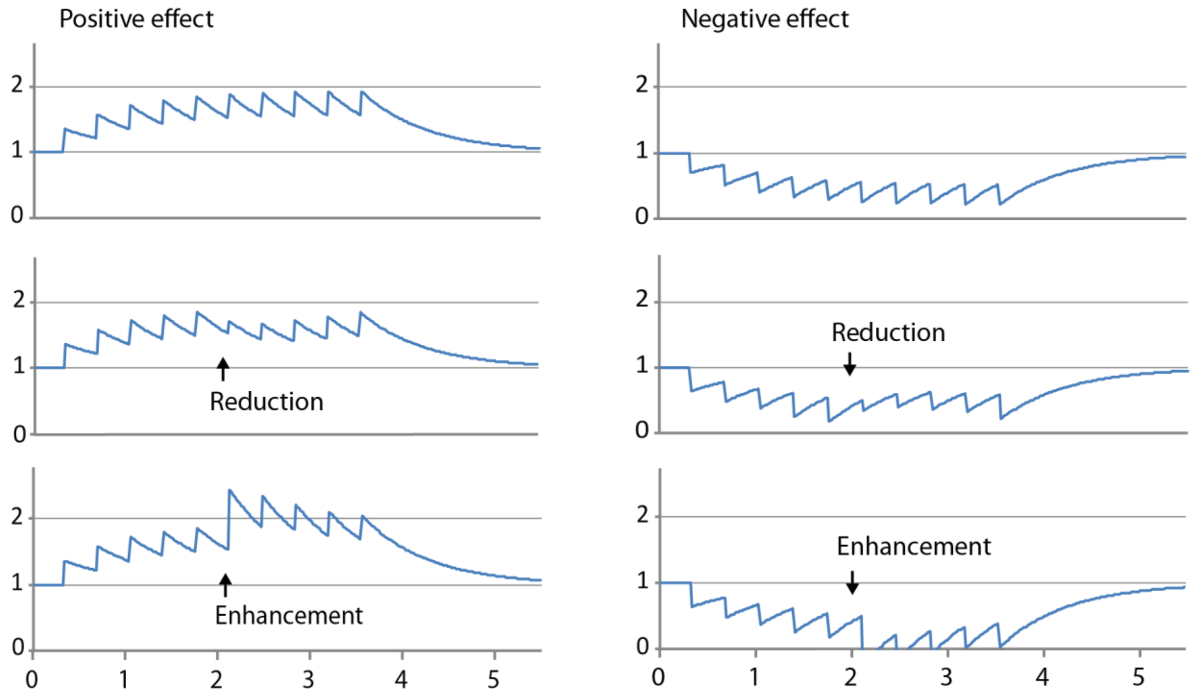

Figure 4. *Time series plots showing the size variation of node B for both positive and negative effects as a result of reduction and enhancement.*

does have some persistence but it decays rapidly. For example, it is about half as big after 0.2 seconds.

$$m = m_0 .\ 0.046^t$$

The temporal profile of the size of node B under conditions of simple causality (both positive and negative) with both reduction and enhancement is shown in Figure 4.

Blocking effects were implemented using the simple disc shown in Figure 2c. This appeared when the VCV pulse from node C reached the junction point of the two curves.

2.3 Other implementation details

The VCVs were implemented as Hermite Spline curves, and the animation had the form of a streak having a sinusoidal profile along its length. The graph underlying the implementation was a form of Petri net, with timings on the edges corresponding to the time of transition of the color pulses. Nodes incorporated the energy function specified above. Edges incorporated the enhancement and reduction functions.

3.0 Study: Causal Interactions Using Multitouch

A study was run to test the claim that simple animations between visual objects can convey a rich causal semantics with an interactive multi touch screen interface. The method was adapted from Wolf (2007). He had participants complete a linguistic description of the observed behavior of the system by inserting words such as cause, enable or prevent into a sentence describing the boat, fans and result.

Two of them involved simple touch, and two involved multi touch.

The conditions were as follows.

1. Increase
2. Decrease
3. Increase Blocking (IB)
4. Decrease Blocking (DB)
5. Increase Enhance (IE)
6. Increase Reduce (IR)
7. Decrease Enhance (DE)
8. Decrease Reduce (DR)

3.1 Method

The study was carried out in a screened off desk in a student union building at the University of New Hampshire. Participant recruitment was by means of a poster affixed to the outer side of the screen. It was administered by a paid undergraduate research assistant. Participants were rewarded by being given a set of pens or a notebook. There were 28 participants who took part in the study. They were judged to be mostly undergraduate students although they were not asked. Participants were told that the experiment was about computer interfaces, and asked to read the IRB consent form and sign if they agreed to participate.

Conditions 1 and 2 were always given first to establish whether the basic effect was perceived, prior to testing interactions. Condition 2 was given before condition 1 for half (randomly determined) of the subjects.

Participants were instructed as follows:

> *Please touch the blue ball on the left to see what happens. What do you think is going on?*

Participants were then asked to circle either *increase* or *decrease* in the following written sentence:

Blue causes green to [increase, decrease].

The remaining 6 conditions were given in a random order and participants were instructed as follows:

> *Please touch the blue and red colored balls on the left to see what happens. Try touching them separately and together. What do you think is going on?*

Participants were asked to circle one of the words in each of the square brackets:

Blue causes green to [increase, decrease].
Red [blocks, enhances, reduces] the effect of Blue on Green.

3.2 Results

The distribution of results is shown in Table 1. The rows represent the effect the design was intended to convey. The columns represent the responses for conditions 3 through 8.

Table 1

	IB	DB	IE	IR	DE	DR
IB	26				1	1
DB		26			1	1
IE	1		26		1	
IR	4		4	20		
DE		3			17	8
DR		6	7			15

A CHI squared test on each of the rows shows that all are highly significantly different from chance ($p < 0.001$).

The values along the diagonal in this table show instances where subject responded as intended in the design. Off diagonal responses deviated from this.

4. Discussion

The results add further support to the idea that there is a way of interactively representing causal relationships that is between the very concrete animation of fans and boats (Wolf) and the abstraction of arrows in a node link diagram. The results suggest a method for interactively representing somewhat complex causal interactions having the form: A causes some change in B (A→B) and C modifies that causal relationship.

As discussed in the rational, the main design problem had to do with timing. In order to show a modifying influence of variable C on A→B it is necessary for the A→B effect to persist perceptually so that users can explore how it is modified by C. The solution presented here is based on an energy model expressed through a form of animated Petri net.

Judgments of the different kinds of causal interactions were not made with equal reliability. Causal blocking (conditions IB and DB) was most reliably judged (by 26 of the 28 participants) and it is worth noting that this is as reliable as Wolf's results with animated objects. The enhancement of a positive effect was equally reliably judged (26 of 28 participants), though the reduction of a positive effect was somewhat less reliably judged (20 of 28 participants). The negative causal effects were the least reliably judged. In condition DE causal enhancement of a negative causal effect was correctly judged by 17 of the participants but it was judged to be enhanced causal reduction 8 times out of 28. We can speculate on the reason for this. The problem may have been due to a language ambiguity in the term reduction. Under this condition a negative effect is enhanced – in other words the effect of variable C is to make B decrease more in size. Possibly the term reduction may have been taken as referring to the change in the size of B and not on the primary A→B effect as intended.

Similar explanations can be constructed for the response pattern in the case of the causal reduction of a negative causal effect condition (condition DR). In this case 15 of the 28 subjects judged the effect correctly. Six of the subjects reported a blocking effect (DB). Given that blocking is simply an extreme example of effect reduction this is understandable. Seven of the subjects reported an enhancement of a positive effect (IE). Since the effect of touching C is actually to make node B bigger, this may partially explain the results.

Can these techniques be applied to more complex causal models? Suppose we have a diagram with ten or twenty nodes in an acyclic directed graph. It is easy to imagine the exploration of local interactions can proceed using the methods developed here. However, it seems unlikely that complex chained interactions can be reliably shown. The problem is fundamental to the representation. The animations developed in this paper use an energy model propagated in a kind of Petri net. But the purpose of the Petri net implementation was to support visual perception of causality, not to provide a modeling tool. Longer chain effects and more complex interactions would not be captured correctly. Also, the problems with timings will be multiplied. Much longer persistence of main effects would be needed for users to interactively discover how several secondary causal factors operate, and the secondary effects would have to persist so that more complex interactions could be explored.

In summary, the methods outlined here may be usefully applied to show simple interactions but are unlikely to be useful in the interactive representation of complex models, except to allow users to locally explore parts of them. Nevertheless, these methods do expand the interactive techniques available for the interactive representation of causal relations.

Acknowledgements

This work has been funded in part from NSF grant 1016607. Discussions with Bill Wright of Oculus Inc were helpful in formulating some of the ideas.

5.0 References

BARSELOU, L.W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences.* 22, 577-660.

BARTRAM, L., AND YAO, M. 2008. Animating Causal Overlays. *Computer Graphics* Forum 27, (3) 751-758.

KADABA, N.R., IRANI, P., AND LEBOE, J. 2009. Analyzing animated representations of complex causal semantics. *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization* 77-84.

Kadaba, N.R., Irani, P., and Leboe, J. 2007. Visualizing causal semantics using animations. *IEEE Transactions on Visualization and Computer Graphics* 13(6) 1254-1261.

LOCKE, J. 1690. An Essay Concerning Human Understanding. Reprinted by *Project Gutenberg*.

NEUFELD E. M., KRISTTORN S. K., GUAN Q., SANSCARTIER M., WARE, C. 2005. Exploring causal influences, *SPIE Conference on Visualization and Data Analysis*, San Jose, January. vol 5669, 52-62.

MOORE J.W., LAGNADO D, DEAL D.C., Haggard P 2009. Feelings of control: contingency determines experience of action. *Cognition* 110:279–283 123

MICHOTTE, A. 1963. *The Perception of Causality*. New York: Basic books.

WARE, C. NEUFELD, E. AND BARTRAM, L. 1999. Visualizing Causal Relations, in Late Breaking Hot Topics *IEEE Information Visualization Symposium*

WOLFF, P. 2007. Representing causation. *Journal of Experimental Psychology: General*, 136, 82 - 111.