DYNAMIC

BACKWARD CAUSATION

W. CLIFTON BEAN

DEPARTMENT OF MATHEMATICAL SCIENCES

UNIVERSITY OF HOUSTON- CLEAR LAKE

First Online Version, March 2008

Comments to the author at wbean002@hotmail.com

ISBN 978-0-9815928-4-8
ISBN 978-0-9815928-5-5

For information or to order additional books, please write:

Tito Publishing Company
P. O. Box 58454
Houston, TX 77258-8454
U. S. A

PROLOGUE

When I published *Naive Causal Modeling* in 1996, I thought its most significant contribution was its argument for a kind of temporal dynamism in which what is real for observers varies with time. My stance in this regard represented an intermediate position between tensed and tenseless theories of time. My plea for an ontological distinction between (1) an underlying reality (Real-World) for which backward-causal events occur at "corners," and (2) a smoothly continuous version of reality (Observer-World) shared by observers, was based upon what I regarded as a repair to existing theories of causation and counterfactuals. That repair was motivated by lengthy consideration of causation in reverse time, and was facilitated by treating cause and effect not as distinct events, but as truth value transitions for different attributes pertinent to a single event. Underpinning the analysis was the premise that causation is most aptly analyzed in terms of discrete time and space.

Unfortunately, in broadcast news parlance, I buried the lead. That is, I placed the discussion of nontraditional backward causation in *Naive Causal Modeling*'s Chapter 10, titled "Toward Backward Causation and Time Reversal Models." Many readers could not get past the first nine chapters, laden with AI jargon, in order to reach Chapter 10's excursion into the philosophy of time.

Accordingly, I have decided to republish my argument for dynamic time as this treatise. Herein, I demonstrate that logical fallacies dominate the philosophical literature on (1) causation under reverse time and (2) backward causation. Proceeding from first principles, I use an intuitive style of presentation accessible to a wide audience. I show that backward causation is ubiquitous in reverse time provided it manifests itself in a dynamic manner profoundly different from standard descriptions. The new formulation of backward causation (1) withstands standard bilking paradoxes and (2) accounts for the phenomenon that, in forward time, causes later than their effects virtually defy detection.

Clearly, in undertaking a fresh analysis of backward causation, I was utterly unmotivated by a desire to retrofit theory to quantum peculiarity. Nevertheless, when I attempted to apply revisionist backward causation to the celebrated quantum mechanical EPR problem, I found that it arguably rendered intuitively plausible the quantitative results predicted by quantum mechanics. At the same time, it eliminated the need to contemplate instantaneous action at a distance. Moreover, a second application seemed to explain the pattern of concentric rings in electron diffraction experiments in such manner as to suggest that backward causation, properly analyzed, is central to conceptual resolution of wave/particle duality.

In closing, I wish to alert the reader to two novel features of *Dynamic Backward Causation*: (1) Occasional use is made of the notion, defended in *Naive Causal Modeling*, that chains of counterfactual necessity can occur interior to an occupied event of spacetime. This presumes that extent is not implicit in the concept of causal propagation. (2) The word "epoch" is used in the traditional sense of "time instant," rather than in the more modern sense of "time interval."

W. Clifton Bean
Houston, Texas
March, 2008

Dynamic Backward Causation

1
CAUSAL MODEL FOUNDATIONS

This chapter implements the premise that causation is most aptly analyzed in terms of discrete time and space. It accomplishes this by treating cause and effect not as distinct events, but as truth value transitions for different attributes pertinent to a single event. It assumes that causation involves both regularity and local counterfactual necessity, yet concedes to regularity theory that many simultaneous interactions display attribute truth value temporal asymmetries useful in discriminating between cause and effect.

1.1 Labeled points and attribute registers

We begin to discuss causal model foundations by building on the intuitive regularity notion that, relative to an intelligence's accumulated database of experience, an event of type A is sufficient for the simultaneous occurrence of an event of type B if invariably whenever A happens, then B happens. Similarly, an event of type A is necessary (in a regularity sense) for the simultaneous occurrence of an event of type B if whenever B happens, then A happens. Events in the real world, however, do not carry just singleton labels. A given occupied point at a given instant (for a given intelligence) may have *multiple attributes*. That is, a given occupied point at a given instant may carry descriptors or truth values for such attributes as color, texture, spatial location, intensity, and orientation. Thus an occupied point is typed not by a singleton label, but rather by an n-tuple of attributes, each carrying a truth value. So an Event at an occupied point may be regarded as the n-tuple's acquisition of an altered truth value array. Indeed, we are free to adopt the change-based notion that a time instant is *generated* by an Event. For example, if (say) n=5, we might have an Event type-labeled by A1=1, A2=0, A3=0, A4=1, A5=1, corresponding to acquisition of the truth value set (1,0,0,1,1) for the ordered attribute array (A1, A2, A3, A4, A5). Here, of course, 0 denotes "false" and 1 denotes "true."

A subset of an Event, associated with the truth value transition of a single attribute, will be called an *event*. This resembles definition by von Wright [91, p. 71] of an elementary event as a change, i.e., from a to ā, "on a pair of occasions which are adjacent either spatially or temporally or both." Although Shoham [79, p. 630] notes that the causal relation holds between "temporal entities–some subset of facts, events, properties, occurrences, actions, factors, and so on," like Bunge [9, p. xix] we intend to restrict attention to relationships involving events.

For attributes that are usually described by a continuum of values (e.g., color/wavelength of reflected light), it is trivial to digitize to a succession of *observations*, an observation of an occupied point at a given instant being a finite set of truth values in an *attribute register*. Color, then, could be represented by a set such as $(\lambda 0 \leq \lambda \leq \lambda 1)$?, $(\lambda 1 < \lambda \leq \lambda 2)$?, … , $(\lambda 6 < \lambda \leq \lambda 7)$? Then, if the correct λ were $(\lambda 1+\lambda 2)/2$, the attribute register reading would be the vector (0, 1, 0, 0, 0, 0, 0). In this regard, an attribute register reading evokes the “state descriptions” of Carnap [11].

Note that, rather than carry one attribute register for each distinct occupied point, it is possible for an intelligence to *merge* attribute registers if desirable (so that not all attributes in a merged attribute register will be defined at the same occupied point), e.g., if it has reason to allow for the possibility of instantaneous action at a spatial distance. It seems natural to superimpose a serial perception, or relational, time model whereby time advances in a locale (i.e., for some attribute register) only via the change of the truth value of one or more attributes in that attribute register. Thus a database for a timespan of observer experience may consist of sequential arrays of observation vectors, one array for each attribute register. Note that the phrase "in a locale" may refer to a particular occupied point in space, but only if the attribute register in question is not a merged one.

In our time model we work with durationless instants (epochs). Time is discrete: Epoch separations are of undefined duration. Time extent is measured as a number of successive epochs that is presumably both countable and finite. This is in the serial perception tradition of (1) Mach [55], who maintains that time is an a posteriori sensation associated with the "work of attention," and of (2) Locke (see [92], p. 123), who conceives time perception as based primarily upon acts of attention to a succession of distinct presentations. It also adheres to the relational-time view of (1) Leibniz (see [18], p. 62), who says, "Time and space are not things, but orders of things," and of (2) Schopenhauer [76], who defines time as the "possibility of opposite states in one and the same thing."

Relational time is consistent with Tucker's [89, p. 89] change-based notion that "events ought to be taken as conceptually prior to instants." Admittedly, it does not mesh with McDermott's [54] dense-time temporal logic, which defines events in terms of instants, nor with Allen's [1] nondense-time scheme, which has events spanning time intervals. Yet it draws on the precedent that others, e.g., Manna and Pnueli [59] and Shoham [80] (in working with infinite causal theories), assume discrete time, a kind of modeling that McDermott [54, p. 112], however, regards as "dangerous."

Regularity analysis with discrete time can be accomplished with classical logic. An appropriate modal logic can be used to accommodate the superimposition of local counterfactual analysis onto regularity theory. A nonmonotonic logic would be required for a dynamic database, with respect to which the statement that an event instantiates a regularity relationship is inherently defeasible. Incidentally, the attribute register approach is compatible not only with change-based, but also with time-based, discrete systems, e.g., by having the output of a two-state digital clock attribute driving the taking of readings.

Shoham and Goyal [81] discuss limitations of change-based systems. It is possible that the attribute register approach repairs certain of these limitations. Thus, although events are indeed durationless, a state occupies a time extent identifiable as an epoch sequence, a discrete analog to a Hayes "episode" [36]. Hence, states can overlap. Naturally, some scheme for the computation of continuously varying quantities such as velocity and acceleration must be appended to any discrete-time model, but this is surely feasible. Shoham [82] attempts to rectify an intuitive advantage he concedes to change-based systems over temporal logics, namely involvement of action in defining future states. Incidentally, regarding primacy of change, Bunge [9, p. 88] finds the concept of relational time enhanced by, but Mehlberg [60, vol. 2, p. 148] finds it incompatible with, relativity physics. Excellent discussions of relational time and space are found in [38] and [84]. Some ideas concerning propagation in relational time and space are given by Bridgman [8].

1.2 Pulses and switches: counterfactual-free causal modeling

Certain interactions that are simultaneous nonetheless involve attribute truth value temporal asymmetries. In fact, a large class of interactions that an intelligence should deem causal (e.g., bat hits ball, ball simultaneously recoils, ball continues to recoil) is made up of uniformities of succession for which the following pattern prevails for every transition (not involving intervention) in its database of a first, agent attribute A=0 to A=1 for which, initially, a second, reagent attribute B=0: There exist neighboring epochs t_0, t_1, t_2 (with $t_0<t_1<t_2$) in association with which (in the absence of intervention) the (A, B) "states" follow the time-ordered sequence (0, 0), (1, 1), (0, 1). Ignoring intermediate epochs at which causally irrelevant attributes change their truth values, we may speak of a succession of (A, B) instants, each new instant characterized by the attainment of a new (A, B) state via the adoption by at least one of A and B of a new truth value. Accordingly, the class of interactions in question is characterized by the causal attribute *pulsing* on via (0, 1, 0) and the effect attribute *switching* on via (0, 1, 1).

Now, like most event-based constructions, the attribute register approach permits the ready identification of temporal asymmetries. Let us connect that fact with a brief elaboration, generally following Bunge [9], of how not just AI practitioners but also philosophers differ as to the essential aspects of causality. For example, Hume [41], Schopenhauer [76], Gasking [30], von Wright [91], Russell [72], Hinckfuss [38], Ducasse [23], Selltiz [78], and many others state that temporal *antecedence* of cause over effect is essential. (This is questioned or denied by Dummett [25], Horwich [40], Reichenbach [70], Chisholm [15], and others.) "Antecedence" postulates that the cause must be prior to, or simultaneous with, the effect. (Russell [73] and Collingwood [16] hold that causes and effects must, furthermore, of necessity be simultaneous.

Moreover, Kant [45] claims that most efficient natural causes are simultaneous with their effects.)

Hume [41], Carnap [10], Reichenbach [70], Hutten [42], Born [6], Selltiz [78], and others state that spatial *contiguity* is an essential aspect of causality. (This is questioned or denied by most empiricists, including Bunge [9], Mach [55], and Pearson [66].) "Contiguity" postulates that cause and effect must be in spatial contact or connected by a chain of intermediate things in contact. As Born [6] points out, even quantum theory, though abandoning determinism, preserves antecedence and contiguity.

Otherwise, most positivists (e.g., [17, 41, 55, 63, 66, 67, 75]) hold that the concept of causation should be replaced by or actually reduced to invariable succession in time. Russell [72], Mach [55], and Reichenbach [70] suggest that the notion of causality is outdated. Russell [72] and McTaggert [58] even venture the view that there is no theoretical difference between causes and effects, that it is arbitrary which one calls which. Yet Ducasse [22, p. 94] notes that in infants the growth of teeth invariably follows, but is not caused by, the growth of hair. Bunge [9, p. 365] states that a pattern of succession in time “may simply be a self-unfolding sequence of states.” Further, he, Mehlberg [60, vol. 1, p. 91], and others observe that if causation were nothing more than uniform succession, special relativity–which establishes that the time order of causally connected, spatially disjoint events cannot vary with observer–could not hold. Terletskii [88] would allow causal anomalies involving superluminary signals, but only if certain observers experience time reversal. In summary, positivists go further than the current author by being willing to jettison as a causal feature not only Mackie’s $necessity_2$ [57, p. 12], which allows a priori inference, but also his $necessity_1$ [57, p. 12], the strong counterfactual sense in which a cause enables its effect.

Turning to AI, we see that Kuipers [48] simulates causality as value propagation with constraints. Like Bunge, Forbus [29] proposes that causality requires some notion of mechanism. Forbus introduces the concept of process as the medium for transfer of causality, and builds upon a nondense-time model due to Allen [1]. DeKleer and Brown [18] adopt a qualitative calculus (also using nondense time) that treats causality as information-passing interactions of device components, and introduce "mythical causality" to order intrastate transitions. Thus, within AI, there are constraint-, process-, and device-centered ontologies, respectively, for the study of causality.

Regarding unresolved disputes concerning causality, there are more: Some treat the relata of causation as facts or sentences rather than events. Some doubt the existence of singular causal sequences. Reichenbach [70] finds the direction of causality derivable from the orientation of irreversible processes; Mackie [57] proposes the opposite derivation. Charniak and McDermott [12, p. 454] treat causality as "a primitive notion, not reducible to simpler concepts." Lewis [51] attempts to reduce local nonsimultaneous causality to chains of counterfactual dependence (using possible worlds to evaluate counterfactuals), but abandons regularity theory in the process–thereby, as it develops, rendering "sufficient cause" undefinable. He argues [52] that counterfactual dependence is time–asymmetric and that the direction of (nonsimultaneous) causation is a consequence. (Lewis' ideas could perhaps be extended to simultaneous causation, but Lewis does not make the attempt.) Horwich [40, p. 158] and Shoham [80, p. 163] perceive the connection the other way round, with causality conceptually prior to counterfactuals. Ginsberg [31, p. 69] finds causality and counterfactuals closely related, but neither reducible to the other. Differences of opinion, even among those willing to speak of counterfactual necessity, as to whether a cause is necessary or sufficient (or both) for its effect are worthy of discussion, but do not require our attention in this treatise.

Relative to a pulse/switch model of simultaneous causality, antecedence would be tantamount to the notion of causal extinction, i.e., effect persistence. That is, the principle would be applied that, just as cause is *generally* prior to effect in nonsimultaneous causality, so cause *generally* manifests temporal priority by extinguishing itself first in simultaneous causality. (As Dummett [25, p. 320] says, "A cause operates on a thing, and once it stops operating, the thing then (i.e., subsequently) goes on in the same way until some further cause operates upon it." Or as McDermott [54, p. 123] says, "Clearly, what events must cause directly is persistences, not the truth of facts.")

Indeed, the fact emerges that a pulse/switch causal model–when built upon an attribute register at an occupied point–*automatically* satisfies requirements of regular succession, direction of causation, and spatial contiguity (indeed spatial coincidence). Accordingly, a pulse/switch formalism is admirably suited to serve as a standalone causal model for those whose concept of causality acknowledges no other features. By so serving, a pulse/switch model would render precision to something most writers leave vague: It would make manifest spatiotemporal contiguity and directedness aspects. That is, a causal model user who *elects* to impose a spatial contiguity requirement may employ attribute register design to automate appropriate limitation of the set of possibly causally involved attributes. Furthermore, preliminary investigation by the author suggests that a pulse/switch formalism is readily extended to the problems of modeling causal spatial propagation and causal intervention.

In that regard, a pulse/switch model aptly captures the fact that simultaneous causality essentially precludes interpolation (though not joint application) of an antagonistic causal agent. Even so, a pseudo-intervention, in which the persistence aspect of an effect attribute's new truth value is distributed, i.e., propagated through other attributes via a causal chain, could be modeled. A chain reaction would be typified by a sequence of causal events, in all but the first of which an effect attribute adopted the role of a causal attribute at an epoch when the attribute otherwise would manifest its persistence aspect. Truth value *reversion* of a causal attribute would be an event for which an attempt to furnish a causal explanation would constitute "confused physics." Rather, such a truth value reversion would be intimately related to what a serially perceptive observer meant by time. (Relational time does not advance until some attribute truth value changes.) Relative to a pulse/switch model, truth value reversion of a causal attribute to its value just prior to its joint transition with an effect attribute would not be an event requiring explanation, but rather would be inherent in the concept of cause.

A methodology that flagged pulse/switch pairings would identify a significant subset of the interactions that should be identified as causal. However, owing to the author's view of the local sine qua non (Mackie's necessity_1) aspect of a cause, "counterfactual-free causal modeling" based upon pulses and switches at present is relegated to the task of flagging joint truth value transitions for subsequent or external causal evaluation.

For example, given the (A, B) truth value time sequence (0, 0), (1, 1), (0, 1) (*possibly* manifesting causal extinction), the temporal asymmetry between A and B would be an excellent marker that an automated intelligence could use to set aside a pattern of such interactions for causal evaluation. Unfortunately, regularity of the given (A, B) truth value time sequence is (in the terminology we commend) neither logically necessary nor logically sufficient for an interaction to be causal. It is not logically necessary, for the seeking of magnetic north by a compass needle is said to be an effect of the temporally symmetric insertion of a compass into the Earth's magnetic field, and not vice versa, notwithstanding the absence of a persistence aspect of either attribute truth value change with respect to the other. It is not logically sufficient, for a methodical watchmaker may begin an enduring process of eating a sandwich every day just as a clock strikes twelve without the striking of the clock *causing* the watchmaker to eat the sandwich.

1.3 Fields, uniformities of coexistence, and uniformities of succession

The absence of that temporal asymmetry which for the appropriate subclass of interactions we label "causal extinction" may occur in one of two ways. The first (and presumably more difficult to analyze causally) way in which temporal asymmetry may fail to be identifiable occurs when an A truth value transition to 1 is necessary and sufficient (for example) in a regularity sense for a B co-pulse to 1, i.e., when (A, B) invariably follows the truth value sequence (0, 0), (1, 1), (0, 0). For example, if A=Moon rising?, and B=Tides coming in?, we may perceive that (A, B) invariably follows the truth value sequence (0, 0), (1, 1), (0, 0). Plainly, such a regularity could be treated as a uniformity of *coexistence*, e.g., $A=1 \overset{co}{\Longleftrightarrow} B=1$, the notation in accordance with the following:

Definition 1.1. For arbitrary two-truth-valued attributes A and B, and for necessity and sufficiency interpreted at least in a regularity sense, we write

(1) A=1 $\xrightarrow{co}$ B=1 to denote A=1 sufficient for the coexistence of B=1.

(2) A=1 $\xleftarrow{co}$ B=1 to denote A=1 necessary for the coexistence of B=1.

(3) A=1 $\xleftrightarrow{co}$ B=1 to denote A=1 necessary and sufficient for the coexistence of B=1.

Some lawlike uniformities of coexistence, according to Hempel [37], are the ideal gas law, the relationship between the length and period of the swing of a pendulum, and Ohm's law. If, like Hempel [37] and Mill [63], we ordinarily require that to be causal a law must provide for a uniformity of *succession*, the question arises, What criterion could lead us to use causal language for the moon/tides type of regularity?

The second way in which temporal asymmetry may fail to be identifiable occurs when there is no discernible pattern regarding which one of two attributes A or B has its truth value extinguished first after simultaneous transition from 0 to 1. For example, if A=Match being struck?, and B=Match flaming?, then the (A, B) sequence (0, 0), (1, 1) may be followed by any one of (0, 0), (1, 0), or (0, 1).

Now a rule for the becoming true of an attribute B may involve not only the simultaneous *becoming* true (not by remaining true) of an attribute A, but also the simultaneous *being* true (possibly by becoming true) of an attribute C. For example, define A=Match being struck?; B=Match flaming?; C=Oxygen in room? Here, we have two uniformities of coexistence (each involving a field, or catalytic, attribute C), namely

$$C=0 \xrightarrow{co} B=0$$

and

$$C=1 \xrightarrow{co} R=1$$

where R is a rule stating "If A(1) and B_0=0, then B(1) and A_0=0." The preceding notation is in accordance with the following:

Definition 1.2. For arbitrary two-truth-valued attribute A, we write

(1) A, simply, to denote the truth value of attribute A at a current epoch.

(2) A_0 to denote the truth value of attribute A at the previous epoch.

(3) A(1) to denote A=1 with A_0=0, i.e., A is true, but was false at the previous epoch, i.e., "A is becoming true."

Of course, for two-truth-valued attributes A and B, we write simply (A,B) to denote a current (A,B) truth value pair, and we write (A_0,B_0) to denote the (A, B) truth value pair at the previous epoch (a more convenient choice than the previous (A, B) epoch).

With regard to the uniformity of coexistence linking C=1 and R=1, note that its right member is itself a uniformity of succession. Note also that, when R is operative, any (A, B) state may follow any other (A, B) state with the single exclusion that (1, 0) may not follow (0, 0). (For example, a match, having flamed, may still be struck, giving us a (1, 0) state following a (1, 1) state. If the match is not completely extinguished, a second flare-up could provide a (0, 1) state immediately thereafter.)

For economy of expression, we adopt the following notation:

Definition 1.3. For any pair of two-truth-valued attributes A and B, we write

(1) $\hat{A}_B$=1 to denote A(1), B_0=0.

Consequently, we write

(2) $\hat{B}_A$=1 to denote B(1), A_0=0.

We apply the definitions of $\hat{A}_B$ and $\hat{B}_A$ in the obvious way when at least one of A or B is a disjunction. For example, $\hat{A}_B$=1 if and only if (1) there exists a disjunctive component AI of A such that AI=1, (2) no disjunctive component of A has initial value 1, and (3) no disjunctive component of B has initial value 1.

An unpursued alternate definition of the symbols $\hat{A}_B$ and $\hat{B}_A$ is : $\hat{A}_B$=1 if and only if A is partly changing to 1 and B_0 is partly 0; $\hat{B}_A$=1 if and only if B is partly changing to 1 and A_0 is partly 0. Here, an attribute is "partly changing to 1" if and only if at least one of its disjunctive components is changing to 1, and an attribute is "partly 0" if and only if at least one of its disjunctive components is 0.

Such an alternate definition could be suggested by the field-free example of a human individual possibly "occupied" by one or more of the agents A1=Bacteria?; A2=Viruses?; A3=Allergens?, and possibly displaying one or both of the symptoms B1=Sore throat?; B2=Runny nose? Where A=A1 or A2 or A3 and B=B1 or B2, it might be the case that $\hat{A}_B(1)$ in the alternate extended sense is sufficient over a database for $\hat{B}_A(1)$ in the alternate extended sense. That is, assuming that the individual displaying no symptoms may begin to display either symptom – and that the individual already displaying one symptom must begin to display the other – upon invasion by a new agent, we could have A partly changing to 1 necessary and sufficient in a regularity sense for B partly changing to 1. This could be true even if we assume for consistency with previous usage (1) that one symptom can vanish as the other begins to be manifested, and (2) A partly changing to 1 does not suffice to *maintain* B=1, given an initial value of 1 for both B1 and B2. Note that the individual, initially still occupied by bacteria but not by either other agent, and already having ceased to display either cold-like symptom, presumably begins to manifest one or both symptoms upon invasion by viruses.

In our chosen formulation, although it may be accurate to regard $\hat{A}_B(1)$ as a maximal disjunction of alternate sufficient conditions for $\hat{B}_A(1)$, it is not accurate to regard A as a disjunction of attributes each of whose becoming true is sufficient for B's becoming true, given B_0=0. Rather, A is a larger (or same size) maximal disjunction of attributes each of whose becoming true *whenever each disjunctive component of A has initial truth value 0* is sufficient for B's becoming true, given $B_0=0$. That is, A is a maximal disjunction of attributes such that the becoming true of the *disjunction* is sufficient for B's becoming true, given $B_0=0$.

Obviously, the set of attributes in such a maximal disjunction subsumes (as a subset) the set of attributes whose becoming true *regardless of the initial truth value of the other attributes in that maximal disjunction* is sufficient for B's becoming true, given $B_0=0$. In turn, a disjunction of the attributes of such a subset subsumes the set of attributes each of whose becoming true (without reference to other initial truth values) is sufficient for B's partly changing to 1, given B_0 partly 0.

2
CAUSATION IN REVERSE TIME

Possible counterfactual accessibility metrics, involving a moving arrow of if-time, are suggested. A preliminary model for backward causation is made available by (1) extending the notion of counterfactual accessibility, and (2) proposing a disposition of the so-called bilking paradoxes. In turn, a coherent analysis - needed for certain AI applications - of causation under time reversal is provided. The ubiquitous idea that cause and effect attributes automatically interchange under time reversal is rebutted.

Time reversed necessary causes are studied for simultaneous, minimally delayed, and substantially delayed reactions. The time reversal of a substantially delayed reaction is shown to be a backward causal sequence.

2.1 Metric counterfactual accessibility

It is our eventual goal to illuminate the kind of causal analysis that would apply with reversed time flow, noting in the process the abundance of confusion on that subject in the philosophical literature. Reverse time considerations are sometimes required in AI for troubleshooting. For example, deKleer [19] runs causal analysis backwards to locate circuit faults.

Time reversal concerns the inverted-order experience of an entire event sequence, not the closely related backward causation, time-will-tell conundrum whether, with positively directed time, an effect may precede its cause. Yet in order to permit a more careful analysis of time reversal, it is important that the possibility of recognizing backward causation be discussed first. Preparatory to that discussion, it will be helpful to clarify counterfactual accessibility further in order to indicate the generalization of that notion that permits us to model backward causation.

Regarding counterfactual accessibility, it is relevant that some writers, e.g., [7], distinguish carefully between *logically* possible world and *physically* possible world. These writers tend to regard it as a self-evident bromide that the laws of physics must apply in any physically possible world. Notice, however, that our failure thus far to issue a flat decree that our counterfactual accessibility relation (seen at once to be both reflexive and symmetric) is necessarily transitive leads us to analyze the aforementioned bromide. We are committed to the following assumptions:

1. It is just their counterfactual force that permits us to regard certain regularities in physics as *laws*.

2. Accordingly, "law of physics" is a derived concept, to be analyzed in terms of counterfactuals.

For a relationship to be not only lawful but also causal, we require that relevant real world regularity patterns must persist in certain truncated possible worlds accessible therefrom. Yet we do not mandate that the regularities must have counterfactual force in these accessible possible worlds. As a matter of fact, it is just whenever the accessibility relation is transitive that their having such force is guaranteed.

Accordingly, we leave the door ajar to the possibility of intransitivity of counterfactual accessibility. The reader appalled at the concept of a physically possible world in which the "laws of physics" do not have lawful force may achieve the appropriate narrowing of the scope of interactions that can be deemed casual, without otherwise doing damage to our program, simply by adjoining a requirement for transitivity of counterfactual accessibility. Indeed, we frequently presuppose retention of the laws of physics in possible worlds under consideration. When we do so, however, we make this additional presupposition explicit.

Our primitive intuition of *counterfactually* accessible by or from the real world is based on our experience of what is *actually* accessed by or from the real world. Regarding the former, the real world itself is a natural reference (if=0) [3, pp.80-82] with respect to which hypothetical excursions are analyzed. Regarding the latter, some kind of local maximum-inertia or zero-action path is a natural reference (if=0) with respect to which actual excursions by or from the real world are examined. This last point of view, however, highlights the possibility that more is accessible to the real world that *is* than to the real world that *ought to be*, i.e., the attributed reality that perception suggests exists. We shall presently consider further the possibility of excessively narrow intuition concerning such accessibility. Meanwhile, however, we intend to explore some rudimentary attempts to quantify counterfactual accessibility.

Now the notion of counterfactual accessibility is not incompatible with a metric approach basing possible world proximity to the real world on extent of similarity. It is more readily compatible with a metric based upon similarity of *effort* that would be required to achieve a possible world rather than upon similarity of *result*. In fact, integration onto metricized accessibility of a moving arrow that traversed both time and if dimensions (a moving "real-now" analogous to the "moving now" of continuous or relational time), by affording an opportunity to speak intelligibly about *mobilely* transworld entities, could be a boon to the alleviation of certain nontrivial problems addressed (e.g., [49]) in the philosophical literature concerning retention of self-identity in counterfactual situations.

A nonworldbound entity exhibits transworld retention of self-identity in a counterfactual situation, i.e., in a possible world accessible to the real world at a given epoch, just in case it has a unique counterpart in that possible world with which it may be *identified*. To avoid the presumption of transworld retention of self-identity is to legitimize such questions as "In counterfactual situations, would attribute A continue to be attribute A?" So long as we assume that the past is invariant, we may beneficially adopt the following conventions to confront the transworld self-identity problem.

1. To specify a *worldbound* entity's self-identity at an epoch is to specify its history with respect to the time axis at that epoch in that world.

2. Selection by (or on behalf of) a real world entity of a counterfactual path at an epoch, inasmuch as it generates (dare we say ipso facto?) a possible world different from the real world, yields in that possible world some number (ordinarily one, but perhaps zero or many) of worldbound *replacement* entities (counterparts), not one of which, however, is identified with its worldbound real world counterpart.

3. Counterparts can be "picked out" by the stipulative device of restricting attention to possible worlds that are admissible just because they are not only (1) accessible (in a sense to be elaborated further) to an entity at a real world epoch (a "branch point") but also (2) *identical* with the real

world at all epochs prior to the branch point. Therefore, a real world entity's worldbound counterpart-at-the-branch-point, in a possible world that is admissible at the branch point, is any occupant of that possible world whose history at the branch point may be identified with the real world entity's history at the branch point.

4. A worldbound real world entity's counterpart (in a possible world admissible at a branch point) at a given epoch (in the possible world) later than the branch point is any contemporaneous (with that later epoch) worldbound *successor* within that possible world of a counterpart-at-the-branch-point. (That is, one builds upon the presumption of the existence in the possible world of a procedure analogous to the real world technique of "identifying" a real world entity with its younger self.)

5. Relative to a moving arrow approach that admits entities that are nonworldbound in a dynamic, *mobilely* transworld sense, to specify an entity's self-identity at an epoch is to specify a history at that epoch of its trajectory with respect to both time and if axes. (A transworld entity, then, at a branch point can *become* a counterpart.)

To illustrate, we suppose the entity in question is a particle, and we assume (for expository convenience) continuity of space and time. Then we utilize one spatial dimension x, with x measurements referenced to our certain real world (if=0) particle. (In another context, i.e., when we address hypothetical alteration of the past at an epoch, we find it more natural to let if=0, x=0 designate a hypothetical maximum-inertia or zero-intervention path relative to which a real world particle follows a moving real-now arrow along a trajectory in the (t,x) plane. When we adopt that viewpoint, of course, we presuppose a generalization of our set of five conventions that allows the reference world, i.e., the world with respect to which "relatively counterfactual" excursions at branch points are examined, to differ from the real world. At present, however, it suits our purpose to let if=0, x=0 refer to the real world.) It follows that what we may call the spacetimeif location (x,t,if) of our real world particle at any time t* is simply (0,t*,0), i.e., it is determined exclusively by the moving arrow of time. Yet given an appropriate if metric, the real world trajectory of the particle could perhaps be contrasted to that of certain of its time t_0 counterparts as in Figure 6.

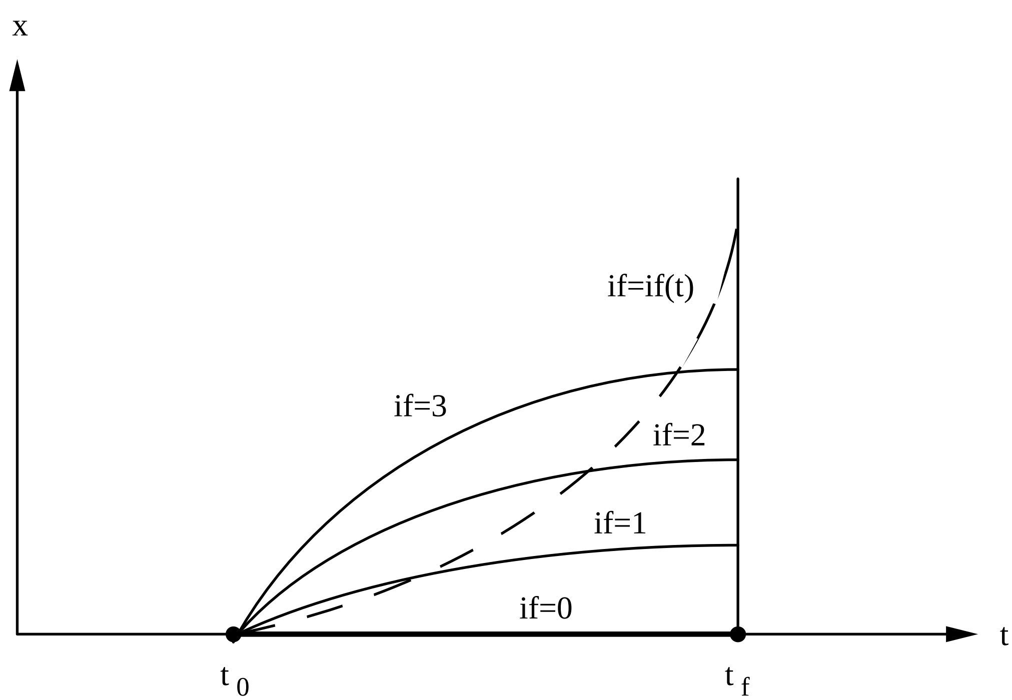

Figure 1: Trajectories of certain counterparts accessible at t_0 to a real world particle

One type of if metric could conform to the scenario that having taken action 1 at t_0 would have transferred the particle onto path if=1 (transformed the particle into its counterpart 1 which, for $t_0 \leq t < t_f$, followed path if=1), taking more difficult action 2 at t_0 would have transferred the particle onto path if=2 (transformed the particle into its counterpart 2 which, for $t_0 \le t < t_f$, followed path if=2), and taking still more difficult action 3 at t_0 would have transferred the particle onto path if=3 (transformed the particle into its counterpart 3 which, for $t_0 \le t < t_f$, followed path if=3). Yet we do not necessarily wish to require that the trajectories of counterparts are all mutually nonintersecting. For example, we do not wish to exclude an if metricization such that a counterfactual action at t_0 could have transferred the particle onto a path whose if value is not constant over the time interval $t_0 \le t < t_f$. Thus perhaps the particle could at t_0 have become a counterpart given by if=if(t) such that if(t) intersects each of the trajectories if=1, if=2, and if=3 between $t=t_0$ and $t=t_f$. (One imagines that if two counterparts momentarily share the same coordinates (x,t,if) they simply *merge* at that moment rather than "collide.") One might say that the real world particle could at t_0 have become a counterpart that subsequently was counterpart 1 at $t=t_1$, was counterpart 2 at $t=t_2$, and was counterpart 3 at $t=t_3$. (A trajectory for which if=2, say, only over some extended *subinterval* of $t_0 \le t < t_f$ would seem to be excluded in a deterministic domain in the absence of some additional *action* between t_0 and t_f.)

For simplicity, the entities we now introduce are generalized particles. A method for generalizing certain more complicated entities, i.e., those involving a state vector based upon more than a singleton attribute, can be inferred. It is convenient to label as a *simple pseudoparticle* - defined at t_f and also associated with the earlier epoch t_0 - any counterpart that a real world particle might have been at t_f had the "moving arrow of if-time" followed over $t_0 \le t < t_f$ an admissible alternate path, counterfactually accessible at t_0. Note that to specify a simple pseudoparticle at t_f is to specify a *history*, i.e., the set of all (x(t, if(t)), t, if(t)) such that $t_0 \le t < t_f$. We shall label as a *simple hyperparticle* - defined at t_f and also associated with the earlier epoch t_0 - the aggregation of a particle that existed in the real world over $t_0 \leq t < t_f$ together with all its counterpart simple pseudoparticles. Here, a simple hyperparticle might be specified by definition over $t_0 \leq t < t_f$ of (x (t, if) , t , if) with t and if treated as independent variables.

It is convenient to label as a *complex pseudoparticle* - defined at t_f and also associated with the earlier epoch t_0 - any counterpart that a real world particle might have been at t_f given the transfer at one or more intermediate epochs t*, $t_0 < t^* < t_f$, from a first counterpart to a counterpart's counterpart counterfactually accessible to the first counterpart at t*. We shall label as a *complex hyperparticle* - defined at t_f and also associated with the earlier epoch t_0 - the aggregation of a particle that existed in the real world over $t_0 \leq t < t_f$ together with all its counterpart complex pseudoparticles.

To illustrate, we observe that corresponding to a particle in Earth orbit at t_0 is a complex pseudoparticle - in lunar orbit at t_f - associated with (1) counterfactual translunar insertion at t_0 and (2) subsequent retrofire behind the moon from that counterpart trajectory onto one of its counterparts at some $t^* < t_f$.

With further regard to the self-identity problem, we note that counterfactual statements thus far considered that (1) assert accessibility at an epoch, and (2) are evaluable as true, implicitly presuppose an invariant past, guaranteeing a kind of one-sided (past-oriented) self-identity. Thus there seems to be no insurmountable conceptual difficulty in entertaining the counterfactual "If I had floored the accelerator pedal, I (that is, the counterpart I would have thereby become) would not have had to stop at the traffic light." (A routine practice that ordinarily presents no difficulty is to identify an entity with successive counterparts that emerge uniquely at subsequent epochs in the same world, i.e., to identify younger self with older self. Yet there is likely a greater difference between the real world "I" at age 50 and the real world "I" at age 20 than between the two versions of myself at a given epoch-one who floored the accelerator pedal and one who did not.)

Could a moving arrow of if-time continue to advance as time stood still? One approach to a metricized if dimension would involve the traversal, at an epoch, of attribute register readings that might have been obtained via *successive* counterfactual alterations of a set of control attributes [3, pp. 52-71]. However, ordinarily such possible worlds are all mutually accessible and do not form an ordered set based upon any substantial inherent distinguishability. In Section 3.1 we discover a situation that more plausibly suggests a moving arrow of if-time that can advance when time does not.

Another way to superimpose a metric upon a counterfactual truth evaluation method based on accessibility via control attribute truth value permutation is via quantification of the notion "would have been accessible." Such quantification would lead to the idea that the nearest neighboring worlds (those with accessibility of order zero) are those worlds that *could have been achieved* at that epoch via counterfactual input of a (presumably different) array of *control attribute truth values.* (Tangential issues concern (1) clarification that any world, real or possible, is accessible from itself and (2) resolution whether or not different worlds could emerge from input of identical counterfactual control attribute truth value vectors at an epoch.) The next nearest neighboring worlds (those with accessibility of order one) at an epoch are those worlds that *could have been accessible* at the epoch via counterfactual input of a different control attribute truth value array at *one* previous epoch. Possible worlds with accessibility of order N (N=0,1,2,...) are those accessible only via counterfactual input of different control attribute truth value arrays at N previous epochs.

Indeed, some counterfactual statements applied to an epoch are evaluable as true only relative to an *altered* past, i.e., they refer to a possible world not accessible from the real world at that epoch. Thus the counterfactual "If I had behaved otherwise at a certain epoch in 1985, I could have then completed this treatise" is manifestly false in the real world. However, the counterfactual "If I had performed an element of work at a certain epoch in 1985, I could have completed this treatise then (if, furthermore, I had also performed an element of work at a discrete number of epochs previous to that certain epoch)" should probably be entertained. Here, the elimination of guaranteed one-sided self-identity that ensues when an altered past in 1985 is presupposed suggests the greater importance of careful reformulation of the counterfactual as a statement of accessibility available to one of my *counterparts*, namely the counterpart that I would have evolved into in 1985 by embarking in 1980 (say) upon a different trajectory with respect to time and if axes.

Consider another example: Regarding a complex pseudoparticle in lunar orbit at t*, there is a statement of counterfactual accessibility of order one at epoch t_f applicable to a particle that has been in Earth orbit throughout $t_0 \leq t \leq t_f$, where $t_0 < t^* < t_f$: "If an impulsive retrofire had been performed at t*, insertion into lunar orbit could have been achieved, if at t_0 the particle had been launched into a proper translunar trajectory."

A further example of counterfactual accessibility of order one is the classic statement "If I had bought some ham, I could have had ham and eggs (if, prior to that, I had bought some eggs)."

A possible world obtained by retention of a real world effect at a given epoch but deletion of its simultaneous real world cause is not accessible from the real world. But perhaps it *could have been accessible* (in the sense in which we are using "accessible") at that epoch given a different history (trajectory) in spacetimeif. That is, perhaps it would have been accessible given the actuality of input of different arrays of control attributes at a discrete number of its previous epochs. Indeed, perhaps *no* counterfactual alteration of control attributes at the epoch (from a counterpart trajectory) would have been required to replicate the real world effect (without its real world cause).

Possible worlds sharing the same order of neighborliness may be drastically unlike. Indeed World 1, the possible world in which the nuclear button was pushed at a certain epoch in 1985, was *accessible* at that epoch (its order of neighborliness to the real world was zero), whereas World 2, the possible world in which the author completed this treatise at that epoch in 1985, has a fairly large positive integer for its order of neighborliness to the real world. Yet as measured in terms of similarity of *results*, World 1 would clearly differ from the real world more than World 2.

To illustrate ordinal metricization of the if dimension based on orders of accessibility, and to illuminate control attribute involvement in the distinction between an accessible world and a could-have-been-accessible world, let us introduce a space vehicle with a thruster attached to its body in such a way that the space vehicle's velocity V can only be augmented instantaneously in the direction of motion by a small fixed impulse ΔV. We designate as ϕ the associated control attribute. Thus $\phi=1$ at any epoch at which a velocity increment is being applied, and $\phi=0$ at all other epochs. We designate as V^+ the velocity acquired at an epoch and as V^- the velocity acquired at the previous epoch. Of course $V^+=V^-+\Delta V$ at any epoch for which $\phi=1$.

Now suppose that the space vehicle's velocity at t* in the real world is V_0, there having been four previous unrealized opportunities (at t_1, t_2, t_3, t_4) for a ΔV maneuver. Accordingly, at t*, the real world (which we designate W0) is characterized by $V^-=V_0$, $V^+=V_0$, $\phi=0$. Accessible (neighborly of order zero) to the real world at t* is the possible world $W0_A$ characterized by $V^-=V_0$, $V^+=V_0+\Delta V$, $\phi=1$.

Had we had $\phi=1$ at precisely one of the four previous ΔV opportunities, in place of the real world at t* we would have had the possible world W1, neighborly of order one to the real world and given by $V^-=V_0+\Delta V$, $V^+=V_0+\Delta V$, $\phi=0$. Accessible to world W1 at epoch t* is the possible world $W1_A$ given by $V^-=V_0+\Delta V, V^+=V_0+2\Delta V, \phi=1$. World $W1_A$ *would have been accessible* from the real world at t* via minimal alteration of the real world's past.

It is noteworthy that $\phi=1$ in world $W0_A$ means that neither world W1 nor world $W1_A$ is accessible from world $W0_A$ at epoch t*. (The fact that the ΔV capability is already being expended in world $W0_A$ makes world $W1_A$ inaccessible therefrom.) Continuing in the indicated manner, we generate Table I.

Table I: Certain Possible Worlds for a Space Vehicle Which Can Augment Velocity by a Fixed Increment ΔV

Possible World	Velocity V^- Achieved at Epoch Previous to t*	Velocity V^+ Achieved at Epoch t*	Is Control Attribute ϕ Being Applied at t*?	What is the Order of Neighborliness at t* to the Real World W0?
W0	V_0	V_0	0	0
$W0_A$	V_0	$V_0+\Delta V$	1	0
W1	$V_0+\Delta V$	$V_0+\Delta V$	0	1
$W1_A$	$V_0+\Delta V$	$V_0+2\Delta V$	1	1
W2	$V_0+2\Delta V$	$V_0+2\Delta V$	0	2
$W2_A$	$V_0+2\Delta V$	$V_0+3\Delta V$	1	2
W3	$V_0+3\Delta V$	$V_0+3\Delta V$	0	3
$W3_A$	$V_0+3\Delta V$	$V_0+4\Delta V$	1	3
W4	$V_0+4\Delta V$	$V_0+4\Delta V$	0	4
$W4_A$	$V_0+4\Delta V$	$V_0+5\Delta V$	1	4

In Table I the question "What is the order of neighborliness at t* to the real world W0?" is tantamount to the question "At how many of t_1,t_2,t_3,t_4 does $\phi=1$?" Now a hypothetically achievable time sequence of worlds (corresponding to the ϕ sequence 0,1,0,1,1 at t_1,t_2,t_3,t_4,t^*, respectively) that a space vehicle might counterfactually have traversed at t* is indicated by the sketch in Figure 2.

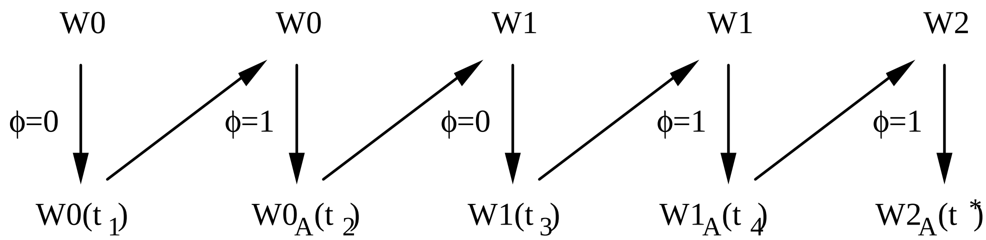

Figure 2: One sequence of possible worlds that might have been traversed at t* by a space vehicle which can augment velocity by a fixed increment ΔV

Incidentally, further illustration can be gained by making different suppositions, namely (1) existence of a comparatively large positive integer N such that $(N+1)\Delta V=V_0$, and (2) replacement of the real world with a world WN such that N velocity impulses ΔV had been accomplished prior to t*. From world WN there would be accessible a world WN_A (neighborly of order N to the real world) in which a velocity increment at t* accomplished a velocity double the real world velocity.

Just as we can embed a relational time that is isomorphic to the integers into a time flow isomorphic to the real numbers, so can we embed the relational if measurement called orders of neighborliness into an if flow isomorphic to the real numbers. Perhaps an example involving continuity everywhere (even during burns) of the first partial derivatives of the space coordinates of a vehicle with respect to time and if will provide a fresh perspective on if-accessibility. Let us introduce a space vehicle with a thruster (1) capable of providing uniform acceleration K during a continuous burn and (2) attached to the body in such a way that the velocity can only be augmented in the direction of motion. Suppose that the space vehicle's velocity at time t in the real world is V_0, there having been a single unused opportunity to initiate a burn at t=0. Here, continuous metricization of the if dimension based on burn duration p is suggested, i.e., *orders* of accessibility are replaced by (continuous) *extent* p of accessibility. In short, if=p. Denoting partial differentiation via subscripts, we can easily construct Table II, depicting the motion of a family of pseudovehicles.

Table II: Motion of Pseudovehicles Initially Counterfactually Accessible to Extent p, the Burn Duration of a Space Vehicle Capable of Uniform Acceleration K

t	x	x_t	x_p	x_{tt}	x_{pp}	$x_{tp}=x_{pt}$
t<0	V_0t	V_0	0	0	0	0
0≤t≤p	$V_0t+\frac{1}{2}Kt^2$	V_0+Kt	0	K	0	0
t>p	$(V_0+Kp)t-\frac{1}{2}Kp^2$	V_0+Kp	Kt-Kp	0	-K	K

Note that in Table IV, given K >0, x_{tt} is discontinuous at t=0 and t=p, and both x_{pp} and x_{tp} are discontinuous at t=p. Even at hypothetical burn initiation and termination, however, the position coordinate x changes continuously with respect to the if parameter p.

2.2 Backward causation and bilking

Our notion of counterfactual accessibility at t_0 developed thus far may treat as an admissible simple pseudoparticle - defined at t_f and also associated with the earlier epoch t_0 - one for which

$$\text{if} = \begin{cases} 0 \text{ for } 0 \le t < t_0 \\ 1 \text{ for } t_0 \le t \le t_f . \end{cases}$$

Yet it excludes as inadmissible as a simple pseudoparticle a hypothetical entity for which if=1 for $0 \le t \le t_f$. Although the simple pseudoparticle and the other hypothetical entity share the same (x,t,if) state $(0,t_0,1)$ at t_0, the differences in their *past* at t_0 make them different individuals in a more profound sense than that in which a particle differs from a pseudoparticle which shares with it at t_0 a common past.

Our admissibility assumption, then, has been that a particle might have behaved otherwise at t_0 and perhaps beyond, but at t_0 could not have changed its past. However, we intend to claim that this idea - that the past at t_0 would necessarily remain invariant had any admissible counterfactual action been undertaken at t_0 - would have to be abandoned to allow coherent analysis of a cause (if any should exist) that occurred later than its effect. In this regard, note that it is generally conceded by philosophers that necessity, as well as sufficiency, can operate retrodictively. That is, specifying which one of two paired nonsimultaneous events is counterfactually necessary for the other does not per se identify a direction of advancing time.

Thus far it has been convenient to adopt a point of view wherein a real world path is labeled if=0 and counterpart paths of various orders of accessibility form a family of trajectories if=if(t). Subsequently, we generally adopt an alternate viewpoint wherein if=0 designates a kind of local maximum-inertia or minimum-intervention path relative to which a real world entity follows a moving real-now arrow along a trajectory in the (t,if) plane.

The view that counterfactual control attribute truth value modification could influence the present and thereby the future is relatively noncontroversial. We have also noted the less common feature of ordinary language that treats as true certain epoch-specific counterfactual statements that tacitly presuppose modification of the past of an entity occupying that epoch, e.g., "If I had sold Polaroid common stock in 1980, I would have made a million dollars (if I had bought an appropriate number of shares when it was a new issue)."

A somewhat related but controversial notion popular in science fiction and motion pictures and discussed by certain philosophers, e.g., Horwich [40], involves modification of the past via the medium of travel *into* the past (whereupon "new" chains of forward causation would be unleashed). Horwich discusses the limited extent to which a discovery by Gödel [33] - of certain solutions of the general relativistic field equations involving closed-loop causal propagation in spacetime - provides a physical theory for time travel.

Another controversial notion in philosophy, and one with which we are more directly concerned than time travel, is the possibility of backward causation or "bringing about the past." Here, most writers (e.g., [4,15,25,27,40,65,77]) confine their attention to the possibility of bringing about the past that actually was. Surprisingly, there is no discussion of the possibility of actualization of a counterfactual past, i.e., acquisition of what we might term a "bogus" past.

Now a purported logical refutation of the possibility of a cause later than its effect is the standard "bilking" argument, so labeled by Flew (e.g., [27]), but also discussed by Dummett [25] and others (e.g., [4,15,40,65,77]). That is, one could have acted after the effect had occurred, but before the cause had occurred, in such a way as to prevent the cause from happening. An apparent paradox ensues.

We would like to point out a related kind of bilking that seemingly presents itself where backward causation is concerned. After an event has *not* happened, it should be possible to act in such a way as to bring into unexpected being an unpredestined future cause. Could not this also place reality in a bind?

Let us propose a solution to both these dilemmas, basing it upon the metaphysical principle that any reality worthy of the name has a built-in capacity for conflict resolution, precluding its placement into an irresolvable bind. We propose that introduction of an unpredestined future cause may serve to create a bogus past in which the previous effect has happened.

Now a single moving arrow with both time and if coordinates might make meaningful the term "current reality." Entertainable counterfactuals at an epoch would then be statements of accessibility (regarding how redirection of the moving arrow of if-time could have transformed an entity into a counterpart) applicable to the member of a set of counterparts that actually "is" or "was" at that epoch. Furthermore, although there would of course be a unique past achieved at t_0, the unique past achieved at later epoch t_f might differ from the past achieved at t_0 *even in its elements identifiable as earlier than* t_0. "The past that is" at t_f would not be "the past that was" at t_0.

Accordingly, the term "current" in the expression "current reality" need not tacitly involve a kind of meta-time. Instead, the totality of the past would be a single-valued function of current time. It would be appropriate to label as a "reality change" an acquisition (as would be attributable to imposition of a backward acting cause) by an entity of an alternate past. Of course to alter the past (and therefore the self-identity) of a single entity is to alter the past (and therefore the self-identity) of every entity that shares the same reality. Therefore, any reality change on the macroscopic level would involve an instantaneous transformation of all memories and records.

Let us be relentlessly candid. The Asimovian [2] sense in which we use the label "reality change" makes it neither oxymoronic nor referential to the garden variety change that occurs via selection of a path into the future to which we were not guided by inertia. It is the sense in which the past can vary as a function of current epoch: The expression "current reality" designates the past associated with the current time epoch. Retention of self-identity goes out the door along with the bathwater. For example, a viewpoint that permits the grafting of a new past onto an entity is a viewpoint that could regard the head-on collision of two identical particles as an interchange of their self-identities.

But now the two bilking paradoxes vanish. Waiting until an effect has happened, then acting to prevent its subsequent cause becomes impossible. We approach the causal event along a branch path for which the effect did not happen. If we prevent the cause from happening, no change ensues: the effect simply never happened. (Imagine someone claiming that he acted to prevent the 1970 cause of the 1948 election of President Thomas E. Dewey.) If we allow the cause to happen, a reality change ensues whereby we acquire a bogus past, i.e., (speaking macroscopically) we acquire the artifacts, memories, and written records indicative of having approached the causal event along a branch path on which the effect did happen. There would be precious few clues that such a reality change had occurred.

Furthermore, the presumption that incompatibility with reality is reciprocal allows us to postulate reconstitution of an entity as a counterpart in a possible world. Therefore, perhaps a fate that can befall a particle in the special case when a so-called "bilking bind" necessitates the particle's annihilation from reality is its becoming a counterpart with a past (in the world in which the particle is reconstituted as its counterpart) that is compatible with the bilking action.

The question arises, Who is bilking whom? According to our view, reality continues inexorably onward without having been placed in an irresolvable logical bind. Therefore, we must confront the possibility that it is reality which benignly bilks those who populate it rather than vice versa.

Why risk a drift into paranoia with a theory which itself predicts the virtual nonexistence of empirical data for its own corroboration? The operative word here is "virtual." We are not prepared to claim that reality changes are actually occurring in the realm of macroscopic phenomena, even though we may select macroscopic examples for expository clarity. We hesitate to claim that they occur even at the *measured* quantum level. Perhaps the act of measurement locks in a real world quantum attribute value such that it cannot be altered by subsequent reality changes.

However, in Section 3.1 we provide an interpretation in terms of backward causation of the well known EPR experiments. Therein, we postulate that the apparent operation of exclusion rules for reality changes induced by acts of measurement does furnish a clue that such changes actually occur. Yet whether or not reality changes ever occur to us as we move forward in time, we shall see in Section 2.3 that the concept of reality change is vital wherever there is a need (e.g., in AI applications) for a coherent analysis of causation in reversed time.

2.3 A saw concerning time reversal

We continue by showing how a pulse/switch model has heuristic value in illuminating a saw in the philosophy of science concerning time reversal. In the process we rebut the idea that cause and effect attributes automatically interchange under time reversal. We study, in turn, time reversed necessary causes for simultaneous, minimally delayed, and substantially delayed reactions.

First, we consider the time reversal of a simultaneous causal connection. As we stated in Section 2.1, reverse time flow considerations are sometimes required in AI for troubleshooting, as when deKleer [19] runs causal analysis backwards with modified models and heuristics to identify circuit faults. Unfortunately, deKleer states [19, p. 278] that time flow reversal is the only difference between the two questions "What could have caused x?" and "What does x cause?" This reflects a prevailing view in the philosophy of science that is captured by this quotation from Whitrow [92, p. 323]: "One consequence of a reversal of our sense of before-and-after would be that causes and effects would be interchanged."

Thus, Zwart [94, pp. 91-92] maintains that, if the accustomed order of events were suddenly to change so that those events which we normally regard as causes were to succeed their effects instead of preceding them, we would exchange the terms cause and effect and not the temporal order. Whitrow [92, p. 323] illustrates the point with this example: "Thus, instead of the stone falling to the ground because my hand has released it, the stone would fly upwards in order that my hand should grasp it."

Others, perhaps influenced by a Kantian causal theory of time, fall into the same trap. Thus Terletskii [88, p. 73] uses causal interchange language to illustrate a reverse time sequence of a bullet striking a target. Rothman [71, p. 64] says, "But because the cosmic movie can be run backward in Newton's universe, causes can be transformed into effects and vice versa."

Yet such a view is coherent relative to neither the author's intuition nor the model recommended in this treatise. The incoherency is especially easy to demonstrate for causal pulse/effect switch, time-asymmetric interactions. Let us utilize Whitrow's hand/stone illustration, avoiding the use of attributes whose very definition presupposes a direction of time flow. Thus let us introduce H=Hand changing encirclement configuration?; S=Stone in motion? With positive time flow, there invariably exists a succession of epochs t_0,t_1,t_2 (with $t_0<t_1<t_2$) with respect to which (H, S) adopts the truth value sequence (0,0), (1,1), (0,1). As we have adhered to the Mill view that only attribute truth value *transitions* may count as members of a causal pair, the two members of the uniformity of succession that we evaluate to determine if the connection is causal are (1) the attainment of truth value 1 by H at t_1, and (2) the (simultaneous) attainment of truth value 1 by S at t_1. As we concern ourselves with a field within a database for which the requisite regularity pattern is satisfied, and as S=1 (and any $\tilde{S}$ =1 logically implied by S=1, where $\tilde{S}$ is a subattribute [3, p. 38] of S) at t_1 could have been removed crisply [3, pp.62-63]

so as to leave intact H=1 at t_1 but not vice versa, we adjudge that H(1) at t_1 was a necessary *cause* of S(1) at t_1.

With reverse time flow, (H, S) invariably follows a truth value time sequence (0,1), (1,1), (0,0). But then the pair of attribute truth value transitions to be examined for causality of connectedness is (1) H=1 at t_1 and (2) S=0 at the "later" epoch t_0. This uniformity of succession has the counterfactual property that the S transition to 0 (and any $\tilde{S}$ =0 logically implied by S=0, where $\tilde{S}$ is a subattribute of S) at the "later" epoch t_0 could have been removed crisply so as to leave intact H=1 at t_1 (e.g., "ascending" stone bypasses hand changing encirclement configuration and continues ascending) but not vice versa (without hand changing encirclement configuration, ascending stone cannot come to rest).

We have, in other words, the following: Even with time reversal, the H attribute is still the causal attribute and S is still the effect attribute. However, instead of H(1) causal with respect to simultaneous S(1), the interaction becomes H(1) causal with respect to what we term "minimally delayed" S(0).

As with forward time flow, then, every sensible counterfactual analysis that can be performed in the reverse time flow case identifies hand changing encirclement configuration as causal. We assert, furthermore, that (as with forward time flow) there is a *favored* counterfactual analysis. To support this claim, we now give a more thorough presentation of the (at least) four counterfactual analyses available from the vantage point of (H,S) = (0,0) at t_0 to an observer traveling backward in time. We obtain them by reflecting upon the assumption that, just as (0,0), (0,1) is inadmissible as a sequence of distinct (H,S) states in forward time flow, so must the (H,S) sequence (0,1), (0,0) be excluded in reverse time.

Parenthetically, we note that of course there would exist a tempting alternate point of view that rejected counterfactual analysis altogether and proclaimed that H=1 at t_1 was "predestined" at t_2 and could not have been deleted. At this point, it's acronym time: We hope that subsequently SIM, MDR, and SDR will be suggestive of "simultaneous," "minimally delayed reaction," and "substantially delayed reaction," respectively. (For the time being, we apply the acronyms from the forward time point of view.)

Counterfactual Analysis #SIM1. From (H,S) = (0,1) at t_2, if H had been 0 at t_1 (so that H=0 at each of t_2,t_1,t_0), then S would not have changed to 0 at t_0, i.e., S would have been 1 at both t_1 and t_0. (This analysis takes the view that the "ascending" stone would have become a counterpart that continued to "ascend" - spinning its metaphorical wheels, but with a sense of advancing time, in quest of a subsequent hand-grasping event - in the absence of the hand changing encirclement configuration.)

Counterfactual Analysis #SIM2. From (H,S) = (0,1) at t_2, if H=1 at t_1 had been inadmissible, then the stone would have become a counterpart in a world in which time stopped at t_2. (This analysis is based on the assumption that with reverse time flow (0, 1) as well as (0,0) is excluded as the next (H,S) state after (0,1), i.e., the (H,S) state at t_1. That would be the case where truth value transitions of other attributes serve to identify t_1 as the critical epoch at which any hand-grasping event would have to occur. That would establish that the stone had no counterpart that continued to "ascend" as t "advanced" to t_1 and beyond.)

Counterfactual Analysis #SIM3. From (H,S) = (0,1) at t_2, if H had been 0 at t_1 (so that H=0 at each of t_2,t_1,t_0), then the stone would have become a counterpart in a world in which H's becoming 1 at t_1 is not causal with respect to S's becoming 0 at t_0. (This analysis refutes the inadmissibility of the (H,S) reverse time sequence (0,1), (0,0), and hence presupposes that an ordinarily inaccessible world somehow becomes accessible in special situations. Such an analysis undoubtedly would not be *favored* by our time reversed observer.)

Counterfactual Analysis #SIM4. From (H,S) = (0,1) at t_2, if H had been 0 at t_1 (so that H=0 at each of t_2,t_1,t_0), then S=0 at t_0 would have been retained, yet there would have been no transition of S from 1 to 0, i.e., there would have obtained S=0 at both t_1 and t_2. (This analysis amounts to the claim that, at t_1, the S truth value at t_2, i.e., the stone's "past" - and hence the common past of all stone observers - would have been otherwise if H had been 0 at t_1.)

Note that there are also at least four counterfactual analyses that can be performed by a real world stone observer going forward in time. Naturally, all counterfactual analyses involve hypothetical stone removal from the real world (and vice versa). That is why annihilation from reality is not treated as the consequence of a separate possible analysis. Yet counterfactual analyses #SIM1, SIM2, and SIM3 by the time reversed observer go further and postulate the annihilation of the stone from the reality surrogate associated with a counterfactual analysis that an observer moving forward in time (and having experienced all three epochs t_0,t_1,t_2) would confidently *prefer*: "If the hand configuration hadn't changed at t_1, the stone wouldn't have begun to fall at t_1."

Only if we accept counterfactual analysis #SIM4 by the time reversed observer do we find the time reversed stone continuing, in the same alternate reality, to be a reflection of the forward time version of itself. Accordingly, we claim it is the counterfactual analysis that a reversed time observer would *favor*.

How could counterfactual analysis #SIM4, involving as it does an altered past, ever be preferred (even by an observer going backward in time)? First of all, perhaps it ought not surprise us that an altered reality (involving continuation on a maximum-inertia path with intervention removal) that is routinely viewed as accessible to the real world from the forward time point of view would also be so viewed from the reverse time point of view. Next, "acquiring an altered past" in this context may properly be interpreted as "deleting a bogus past."

The time reversed observer may reason in this fashion: "It was the causal event H(1) at t_1 that induced a reality change, awarding me an alternate past in which S was 1 at t_2 so that S might become 0 at t_0. If the causal event H(1) had not occurred, i.e., if H had been 0 at t_1, there would have been no reality change, and I (from the forward time viewpoint, 'a replacement individual in my stead') would have experienced a natural past in which S was already 0 at t_2."

Finally, note that it is typical in reverse time scenarios for a fully specified present state to be consistent with multiple possible "prior" branched paths. In this regard, observe that in the previous paragraph our forward time bias led us to cringe at an apparent abuse of the personal pronoun "I" in the reverse time traveler's counterfactual analyses. Yet we see no advantage in belaboring the point, and indeed we find such attributed usage defensible. Accordingly, in pursuit of minimum distraction, *from now on we simply adopt the indicated liberal ascription of personal pronouns to the reverse time traveler without further comment.*

The forward time traveler's sense of current self-identity generally is associated with (1) an invariant past, (2) multiple possible paths into the future, and (3) the path taken into the future influenced by what happens *at_present.* Subsequent to such a branch point, the counterfactual analysis that the forward time traveler tends to make is to deny the "future" that occurred following the branch point and to affirm the past that occurred previous to it. Thus, from the vantage point of the future, he is likely to favor the assertion "If my present at that branch point had been otherwise, my past would have remained the same, but my path into the future would have been different and I would now be on a different branch." The world of the forward time traveler is a world with causes generally earlier than, or simultaneous with, their effects. It is only reluctantly that such an observer will acknowledge that, given the reality of the present and future (relative to the branch point), his past would have had to be otherwise.

The reverse time traveler's sense of current self-identity must generally be associated with (1) an invariant future, (2) multiple possible paths from the past, and (3) the path taken from the past influenced by what happens at present. Subsequent to such a branch point, the counterfactual analysis that the reverse time traveler must tend to make is to affirm the "future" that occurred following the branch point but to deny the past that occurred previous to it. Thus, from the vantage point of his future, he is likely to favor the assertion "If my present at that branch point had been otherwise, my subsequent future would have remained the same, but my path from the past would have been different and I would have arrived at that present along a different branch." The world of the reverse time traveler is a world (as we shall see) with causes generally later than, or simultaneous with, their effects. It is only reluctantly that such an observer will acknowledge that, given the reality of the present and past (relative to the branch point), his future would have had to be otherwise.

The forward time traveler tends to recognize the existence of multiple counterparts that he could become in his future, the actualized counterpart dependent upon the present. The reverse time traveler tends to recognize the existence of multiple counterparts that he could have been in the past, the actualized counterpart dependent upon the present.

The forward time traveler's precognitive impressions are ambiguous regarding by what path he will depart from the present, with the nature of the here and now perhaps resolving that ambiguity and in any case channeling him from an invariant past. The reverse time traveler's "memory" is ambiguous regarding by what path he has arrived at the present, with the nature of the here and how perhaps resolving that ambiguity and in any case funneling him into a predestined future.

Yet past alteration is not forced upon us in the analysis of the time reversal of a simultaneous causal interaction. Next to counterfactual analysis #SIM4, the no-change situation of counterfactual analysis #SIM1 (which doesn't involve an altered past) is perhaps most similar to the standard forward time counterfactual analysis. Furthermore, it is a viable analysis where buying time awaiting a causal event would be admissible. What *is* forced upon us in each of the four reverse time analyses is the removal of an S switch to 0 as a consequence of the counterfactual removal of an H pulse to 1. There is no reciprocal effectivity [3, pp.36-42]: the H pulse to 1 is causal in reverse time.

Before we present examples of time reversed causal minimally delayed reactions and substantially delayed reactions, we state (from the pulse/switch viewpoint) some definitions. A causal interaction is said to be "minimally delayed" if and only if the truth value reversion (say to 0)-of the agent attribute whose truth value transition (say to 1) is the causal event-*coincides* with the reagent attribute truth value transition that is the effect event. For example, where an A transition to 1 is causal with respect to a B transition to 1 in the (A,

B) sequence (0, 0), (1, 0), (0, 1), the causal interaction is minimally delayed. A causal interaction is said to be "substantially delayed" if and only if the truth value reversion (say to 0)-of the agent attribute whose truth value transition (say to 1) is the causal event-*precedes* the reagent attribute truth value transition that is the effect event. For example, where an A transition to 1 is causal with respect to a B transition to 1 in the (A, B) sequence (0, 0), (1, 0), (0, 0), (0, 1), the causal interaction is substantially delayed.

Next, we consider the time reversal of a causal minimally delayed reaction. An example of such an interaction, given positive time flow, involves the attributes W=Wall switch being pulsed?; L=Overhead light on? Given a positive time sense, we assume there invariably exists a sequence of epochs t_0, t_1, t_2 (with $t_0<t_1<t_2$) with respect to which (W, L) adopts the truth value sequence (0, 0), (1, 0), (0, 1). Again, as only attribute truth value *transitions* may count as members of causal pairs, the two members of the uniformity of succession to be evaluated to determine if the connection is causal are (1) the attainment of truth value 1 by W at t_1 and (2) the (minimally delayed) attainment of truth value 1 by L at t_2. As we concern ourselves with a field within a database for which the appropriate regularity pattern is displayed, and as L=1 (and any $\tilde{L}=1$ logically implied by L=1, where $\tilde{L}$ is a subattribute of L) at t_2 could have been removed crisply so as to leave intact W=1 at t_1 but not vice versa, we adjudge that W(1) at t_1 was a necessary *cause* of L(1) at t_2.

With reverse time flow, (W,L) invariably follows a truth value sequence (0, 1), (1, 0), (0, 0). But now the pair of attribute truth value transitions to be examined for causality of connectedness is (1) W(1) at t_1 and (2) the (simultaneous) attainment of truth value 0 by L at t_1. This uniformity of succession has the counterfactual property that the L transition to 0 (and any $\tilde{L}=0$ logically implied by L=0, where $\tilde{L}$ is a subattribute of L) at t_1 could have been removed crisply (e.g., electrical wiring deleted) so as to leave intact the W transition to 1 at t_1 but not vice versa (there would be no L(1), e.g., the light might "remain" on awaiting a subsequent wall switch pulse, if the one at t_1 were removed). Therefore, even with time reversal, W is still the causal attribute and L is still the effect attribute. However, instead of W(1) causal with respect to minimally delayed L(1), the interaction becomes W(1) causal with respect to simultaneous L(0).

Four counterfactual analyses available from the vantage point of (W, L) = (0, 0) at t_0 to a being traveling backward in time can be obtained (as was done with the time reversal of a forward time simultaneous causal interaction) by reflecting upon the following assumption: Just as (0, 0), (0, 1) (like (0, 0), (1, 1)) is inadmissible as a sequence of distinct (W, L) states in forward time flow, so must the (W, L) sequence (0, 1), (0, 0) (like (1,1), (0, 0)) be excluded in reverse time. Again, there is a possible view that rejects counterfactuals and claims that (W, L) = (1, 0) at t_1 was "predetermined" at t_2.

Counterfactual Analysis #MDR1. From (W, L) = (0, 1) at t_2, if W had been 0 at t_1 (so that W=0 at each of t_2, t_1, t_0), then L would not have changed to 0 at t_1, i.e., L would have been 1 at both t_1 and t_0. (This analysis takes the view that the light "already on" at t_2 could have become a counterpart that continued to "remain on" – in quest of a subsequent wall switch pulse event and with a sense of advancing time – in the absence of a wall switch pulse at t_1.)

Counterfactual Analysis #MDR2. From (W,L) = (0,1) at t_2, if W=1 at t_1 had been inadmissible, then the light would have become a counterpart in a world in which time stopped at t_2. (This analysis is based on the assumption that with reverse time flow (0,1) as well as (0,0) is excluded as the next (W,L) state after (0,1), i.e., the (W,S) state at t_1. That would be the case where truth value transitions of other attributes serve to identify t_1 as the critical epoch at which any wall switch pulse event would have to occur, establishing that the light has no counterpart that continues to burn as t advances to t_1 and beyond.)

Counterfactual Analysis #MDR3. From (W, L) = (0, 1) at t_2, if W had been 0 at t_1 (so that W=0 at each of t_2, t_1, t_0), then the light would have become a counterpart in a world in which W's becoming 1 at t_1 is not causal with respect to L's becoming 0 at t_1. (This analysis refutes the inadmissibility of the (W, L) reverse time sequence (0, 1), (0, 0), and hence presupposes that an ordinarily inaccessible world somehow becomes accessible in special situations. Such an analysis undoubtedly would not be *favored* by our time reversed observer.)

Counterfactual Analysis #MDR4. From (W, L) = (0, 1) at t_2, if W had been 0 at t_1 (so that W=0 at each of t_2, t_1, t_0), then L=0 at t_1 would have been retained, yet there would have been no transition of L from 1 to 0, i.e., there would have obtained L=0 at both t_1 and t_2. (This analysis amounts to the claim that, at t_1, the L truth value at t_2, i.e., the light's "past" – and hence the common past of all light observers – would have been otherwise if W had been 0 at t_1.)

As we did for counterfactual analysis #SIM4, we claim that counterfactual analysis #MDR4, involving an altered past, would be favored by a reverse time observer.

The time reversed observer may reason in this fashion: "It was the causal event W(1) at t_1 that induced a reality change, awarding me an alternate past in which L was 1 at t_2 so that L might become 0 at t_1. If the causal event W(1) had not occurred, i.e., if W had been 0 at t_1, there would have been no reality change, and I would have experienced a natural past in which L was already 0 at t_2."

Yet once again, but now for time reversal of minimally delayed forward time causal interactions, past alteration is not forced upon us. The next most palatable analysis available to a reverse time traveler is the no-change view of counterfactual analysis #MDR1, viable where buying time awaiting a causal event would be admissible. What *is* forced upon us in each of the four counterfactual analyses is the removal of an L switch to 0 as a consequence of the counterfactual removal of a wall switch pulse to 1. There is no reciprocal effectivity: the W pulse to 1 is causal in reverse time.

Other examples seem to follow the same pattern. Consider: Every time Hank Aaron hit a home run, the ball transferred from path P1 (pitcher-bat) to path P2 (bat-outfield). With time reversal, one member of the uniformity of succession now in question is the collision of the ball with Aaron's bat, but the other member is the (minimally delayed) transition of the ball from path P2 to path P1. But the transition of the ball from path P2 to path P1 (and any subattribute's truth value acquisition logically implied by the fact of that transition) could have counterfactually been removed crisply while leaving intact the bat hitting the ball (e.g., we *could* be viewing the time reversal of a batting practice situation where Aaron hits a ball coming in from the outfield back along the same path P2), but not vice versa (without bat-ball collision, there would be no transfer to bat-pitcher path P1). With reverse time, Aaron's bat's colliding with the ball is *still* the causal event, but now the effect is the minimally delayed transfer of the ball onto a path from bat to pitcher.

Summarizing our results thus far, and applying inductive reasoning, we may claim for arbitrary time-asymmetrically coupled attributes P and Q the following: (1) The time reversal of P(1) causing simultaneous Q(1) is P(1) causing the minimally delayed reaction Q(0). (2) The time reversal of P(1) causing minimally delayed Q(1) is P(1) causing simultaneous Q(0). (For convenience, we have restricted attention to necessary causes throughout.) Even more succinctly: The time reversal of an SIM is an MDR and vice versa.

It is noteworthy, given the controversy in the literature regarding whether or not an effect may precede its cause, that time reversal of simultaneous (or minimally delayed) causal interactions – even though it does *not* reverse cause and effect attributes – nevertheless preserves causal temporal priority with respect to effect by producing only situations wherein effects occur simultaneously with, or later than, their causes.

Finally, we consider the time reversal of a causal substantially delayed reaction. We soon discover that a different situation obtains regarding preservability of causal temporal priority. For example, collision of a photon packet with a photosynthetic plant may be said to cause a subsequent emittal by the plant of an oxygen molecule. Let P = Photon packet colliding with plant?; M = Oxygen molecule at surface of, or external to, plant? Assume, given a positive time sense, that there invariably exists a sequence of epochs t_0, t_1, t_2, t_3 (with $t_0<t_1<t_2<t_3$) with respect to which (P, M) adopts the truth value sequence (0, 0), (1, 0), (0, 0), (0, 1). Now the two members of the uniformity of succession to be evaluated to determine if the connection is causal are (1) the attainment of truth value 1 by P at t_1 and (2) the (substantially delayed) attainment of truth value 1 by M at t_3. As we concern ourselves with a field within a database for which the appropriate regularity pattern is displayed, and as M(1) (and any subattribute's truth logically implied by the fact of that transition) at t_3 could have been counterfactually removed crisply, leaving P(1)

at t_1, but not vice versa, we adjudge that P(1) at t_1 was a necessary *cause* of M(1) at t_3.

With reverse time flow, (P, M) invariably follows a (t_3, t_2, t_1, t_0) truth value sequence (0, 1), (0, 0), (1, 0), (0, 0). But now the pair of attribute truth value transitions to be examined for causality of connectedness is (1) P(1) at t_1 and (2) the "earlier" attainment of truth value 0 by M at t_2. This reverse time uniformity of succession has the property that the M transition to 0 (and the falsity of any subattribute of M logically implied by M=0) could be removed counterfactually (e.g., via annihilation of the plant's photosynthetic processing apparatus for "absorbing" oxygen molecules) without perturbing the P transition to 1. Yet any reverse time counterfactual sequence for which the P transition to 1 can be removed is a counterfactual sequence for which the M transition to 0 never existed. Keeping in mind that as a cause is counterfactually necessary for its effect so must an effect be counterfactually sufficient for its cause, we conclude that, in the reverse time sequence, the rendering inevitable of "subsequent" P(1) by "prior" M(0) (alternately, the counterfactual necessity of P(1) for M(0)) ensures that it is the pulse to 1 of attribute P that is the causal event. Therefore, there is no interchange of causal and effect attributes with time reversal. Instead of P(1) causal with respect to substantially delayed M(1), the time reversed interaction becomes P(1) causal with respect to *previous* M(0), i.e., a photon packet impingement on a plant causes the prior "absorption" of an oxygen molecule by the plant. In other words, time reversal of a substantially delayed reaction (SDR) would produce a backward causal sequence (BCS), and presumably vice versa. Under such circumstances, causal temporal priority with respect to effect is discarded as a requirement.

A more detailed presentation of the sensible counterfactual analyses that can be performed from the vantage point of (P, M) = (0, 0) at t_0 in the reverse time flow case can be obtained by reflecting upon the assumption that, just as (0, 1), (0, 0), (0, 1) and (0, 1), (0, 0), (1, 1) are inadmissible sequences of distinct (P, M) states in forward time flow, so must the (P, M) sequences (0, 1), (0, 0), (0, 1) and (1, 1), (0, 0), (0, 1) be excluded in reverse time. (As usual, the point of view would exist that rejected counterfactual analysis by claiming that P=1 at t_1 was already inevitable at t_3.)

Counterfactual Analysis #SDR1. From (P, M) = (0, 1) at t_3 and (P, M) = (0, 0) at t_2, if P had been 0 at t_1 (so that P=0 at each of t_3, t_2, t_1, t_0), then M would have remained 0 at both t_1 and t_0. (This analysis takes the view that the oxygen molecule already interior to the plant at t_2 could have become a counterpart that continued to "burrow" into the plant – in quest of a photon "ejection" by the plant – with a sense of advancing time.)

Note: This analysis is not viable for the chosen example. An oxygen molecule could not continue delving into a plant – while observer time continued to advance ad infinitum – in the absence of a photon event. That is, the forward time sequence (0, 0), (0, 0), (0, 1) would be inadmissible in the chosen example.

Counterfactual Analysis #SDR2. From (P, M) = (0, 1) at t_3 and (P, M) = (0, 0) at t_2, if P had been 0 at t_1 (so that P=0 at each of t_3, t_2, t_1, t_0), then the oxygen molecule would have become a counterpart in a world in which time stopped at t_2. (This analysis is based on the assumption that with reverse time flow (0, 0) as well as (0, 1) is excluded as the next (P, M) state after the sequence (0, 1), (0, 0), i.e., the (P, M) state at t_1. That would be the case where truth value transitions of other attributes serve to identify t_1 as the critical epoch at which any photon "ejection" event would have to occur, establishing that the oxygen molecule has no counterpart that continues to "burrow" into the plant as t advances to t_1 and beyond.)

Counterfactual Analysis #SDR3. From (P, M) = (0, 1) at t_3 and (P, M) = (0, 0) at t_2, if P had been 0 at t_1 (so that P=0 at each of t_3, t_2, t_1, t_0), then the oxygen molecule would have become a counterpart in a world in which P's becoming 1 at t_1 is not causal with respect to M's becoming 0 at t_2. (This analysis refutes the inadmissibility of the (P, M) reverse time sequence (0, 1), (0, 0), (0, 0) and hence presupposes that an ordinarily inaccessible world somehow becomes accessible in special situations. Again, this kind of analysis is not likely to be favored by a time reversed observer.)

Counterfactual Analysis #SDR4. From (P, M) = (0, 1) at t_3, if P had been 0 at t_1 (so that P=0 at each of t_3, t_2, t_1, t_0), then M would not have changed to 0 at t_2, i.e., M would have been 1 at both t_2 and t_1. (This analysis amounts to the view that the "approaching" oxygen molecule would not yet have begun to penetrate the plant at t_2, but would have spun its figurative wheels with a sense of advancing time until the impending occurrence of P=1 was assured. The analysis claims that, at t_1, the oxygen molecule's so-called "recent past" - the M truth value at t_2 - would have had to be otherwise if P had been 0 at t_1.)

Note: Counterfactual analysis #SDR4 is viable for the chosen example, i.e., the time reversed oxygen molecule could have become a counterpart which postponed until t_0 entry into the plant, but only if the counterpart somehow was assured of an impending photon packet ejection directly "after" t_0.

Counterfactual Analysis #SDR5. From (P, M) = (0, 1) at t_3, if P had been 0 at t_1 (so that P=0 at each of t_3, t_2, t_1, t_0), then M would not have *changed* to 0 at t_2, but rather would have *remained* 0 at t_2. (This analysis amounts to the view that the "approaching" oxygen molecule was internal to the plant at t_3, t_2, t_1, and t_0. The analysis amounts to the claim that, at t_1, the oxygen molecule's so-called "remote past" - the M truth value at t_3 - would have been otherwise if P had been 0 at t_1.)

Note: Counterfactual analysis #SDR5 is viable for the chosen example and indeed matches the counterfactual analysis that a forward time observer (momentarily coincident with the reverse time observer at t_1) would confidently *prefer* upon experiencing the final of the four epochs t_0, t_1, t_2, t_3: "If the plant had not had a photon packet event at t_1, then there would have been no transition of its surface by an oxygen molecule at t_3." Accordingly, we claim that counterfactual analysis #SDR5, involving alteration of the *remote* past, would be favored by a reverse time observer confronted with a cause that is (for him) later than its effect.

The time reversed observer may reason in this fashion: "It was the causal event P(1) at t_1 that induced a reality change, awarding me an alternate remote past in which M was 1 at t_3 so that M might become 0 at t_2. If the causal event P(1) had not occurred, i.e., if P had been 0 at t_1, there would have been no reality change, and I would have experienced a natural past in which M was already 0 at t_3."

Similarly, an observer moving backward in time, encountering first an unshattering-to-become-a-window (effect) and then "later" an aircraft transition from supersonic (cause), could not delete the aircraft event crisply with respect to the window event without deleting the window event. If there could be a time reversed observer "past" the window event, the aircraft event would be uncontrollably inevitable in the following sense: Its counterfactual deletion (according to a preferred analysis) would transfer the reverse time observer to a possible world where the window event had never happened.

By analogy, we conjecture that a forward time observer's very success in identifying and applying any post-effect cause would eradicate the memory of the reason for doing so. This fact would encourage retention of the illusion of an invariant past.

Finally, referring to Figure 3, we can obtain a similar analysis for a substantially delayed reaction involving a space vehicle initially in Earth orbit.

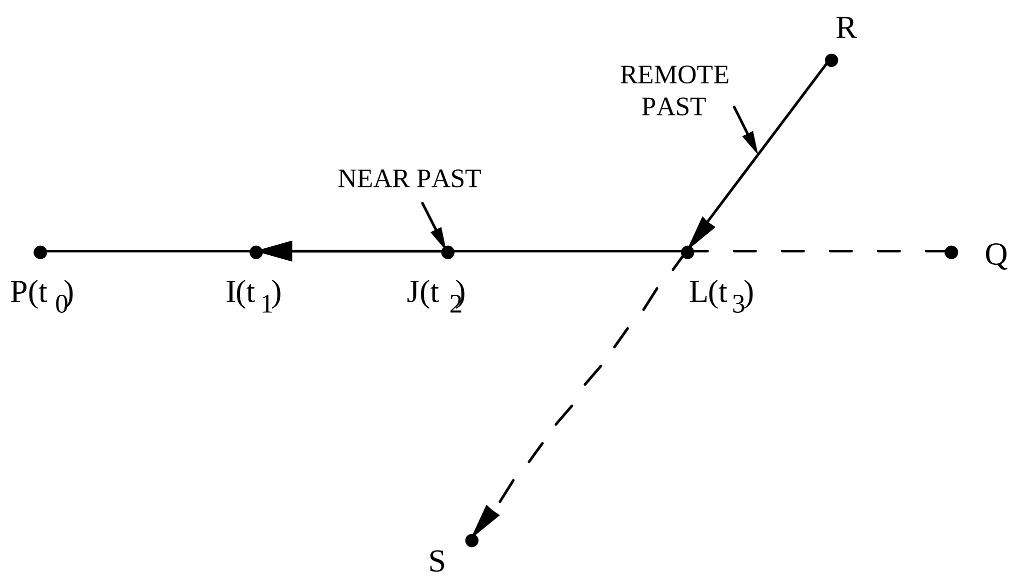

Figure 3: Time reversal of a substantially delayed reaction involving a space vehicle initially in Earth orbit

Suppose I is the pulse initiating an ignition sequence, and L is a launch into an interplanetary trajectory. Since L did occur, the vehicle follows path PILR. If I had not occurred, the vehicle would have continued moving with an orbiting launch platform along path PQ.

What is the reverse time point of view? The vehicle travels along path RL, and then traverses path LI. The I- event is counterfactually necessary for the L-event. The preferred counterfactual analysis (analogous to #SDR5) is that introduction of the I-event actualizes a remote past involving path RL such that the vehicle traveled along path RL, then along path LI. If there had been no I-event, a natural past involving travel along path QLI would have been unreplaced.

Analyses analogous to counterfactual analyses #SDR1 and #SDR4 are not viable. With regard to the former, no reverse time space vehicle counterpart that transfers from an incoming interplanetary trajectory to Earth orbit can delay while other related attributes change truth values: A de-ignition event must occur specifically at t_1. With regard to the latter, no reverse time space vehicle counterpart can "push back to t_0" a transfer into Earth orbit where related attribute truth values change: The insertion event must be at t_3.

That is, it seems dubious that counterfactual detection at t_1 of the I-event would induce a reality change that altered *just* the space vehicle's near past while preserving its remote past. That would involve at t_1 an instantaneous so-called "hyperspatial" jump from spatial position I "before" the reality change to spatial position S "after" the reality change (so that there would have been no transfer from path RL to path LI at t_3). The preferred counterfactual analysis, on the other hand, preserves spatial position at t_1 across a reality change and agrees with a retrospective counterfactual analysis that would be offered confidently by a forward time observer.

A deeper metaphysical analysis of purported hyperspatial jumps in reverse time seems warranted. Unfortunately (the pun is inevitable), we have neither time nor space. More pressing questions arise: By what method would we recognize a cause later than its effect in forward time? Are extraordinary reality changes ever associated with real world departure from some maximum-inertia, action-free reference configuration? Can such extraordinary accessing of a counterpart by the real world ever be actual rather than counterfactual? In Section 3.1 we address these questions in the process of providing an interpretation in terms of backward causation of the EPR experiments. Meanwhile, by invoking our fourth acronym (namely BCS for "backward causal sequence"), we can summarize succinctly the results of this section:

Conjecture 2.1. The time reversal of an SIM is an MDR, and vice versa.

Conjecture 2.2. The time reversal of an SDR is a BCS, and vice versa.

3
DYNAMIC BACKWARD CAUSATION IN FORWARD TIME

An interpretation in terms of backward causation of the notorious EPR experiments renders intuitively transparent violation of the Bell inequality, arguably in just the quantitative manner predicted by quantum mechanics. A new look at nonlocality is afforded. It is argued that the book’s novel treatment of backward causation, by highlighting logical and empirical evidence for the nonmonotonicity of time, should stimulate further research in theories of knowledge and belief. As a parting bonus, a preliminary framework is offered for an appropriately structured new theory of knowledge.

3.1 The EPR experiments

We now consider the experiments in quantum mechanics inspired by a paper (the so-called "EPR paper") by Einstein, Podolsky, and Rosen [24]. The results of these so-called "EPR" experiments seem to involve the acquisition by a particle in one region of space of a definite value of a property by virtue of a measurement carried out on a correlated particle in another, comparatively remote region of space. According to Mermin [61,p.38], such interpretations involve what Einstein called "spooky actions at a distance." The existence of such interaction is inferred from the fact that data from EPR experiments typically violate Bell's theorem in a manner predicted quantitatively by quantum mechanics. Bell's theorem (see, e.g., [20] or [61]) involves an inequality whose derivation reflects the presumption with regard to random data sets that exclusion of a subset believed to be perfectly correlated ought to yield a randomly constituted *remainder* set.

Mermin [61,p.44] describes a hypothetical EPR experiment in which a source produces two correlated particles of spin 1/2 in the singlet state, flying apart toward detectors A and B. "Each detector contains a Stern-Gerlach magnet, oriented along one of three directions..., perpendicular to the line of flight of the particles, and separated by 120˚..." At each detector random determination of the position of a three-position switch governs which orientation is used. "The light on one detector flashes red or green, depending on whether the particle is deflected toward the north (spin up) or south (spin down) pole of the magnet as it passes between them; the other detector uses the opposite color convention."

Two features of typical data from a large number of runs would violate the Bell inequality by reflecting the following properties:

1. Examination of just those runs in which the switches have the same setting reveals that the lights always flash the same colors.

2. Examination of all runs, without regard to switch settings, reveals that half the time the lights flash the same colors and half the time they flash different colors.

These results seem to rule out the possibility (virtually demanded by local realist theory) that each particle bears an instruction set when emitted by the source, fixing in advance the color it will flash for each possible switch setting of its detector. (Thus RGR might denote the instruction set that would lead a detector to flash red for switch positions 1 and 3, but green for switch position 2.) Accordingly, a majority of particle physicists favors alternative explanations in which a detector's measurement itself is causal with respect to the acquisition by a particle of a quantum mechanical property. However, that seems to involve as a consequence the immediate acquisition of the same property by the now distant correlated particle, i.e., instantaneous action at a distance.

We find advantageous an interpretation involving (coherently analyzed) backward causation. We dispute the truth of the counterfactual "If the result of the test had been positive, the organism would have been rodlike" on the grounds that whether an organism is rodlike or not must be causally independent of the result of a measurement. On the other hand, a counterfactual of the form "If the particle had been measured to have positive (rather than negative) spin component along axis one, it would have *been* a particle with positive (rather than negative) spin component along axis one" is very likely evaluable as true in the realm of quantum mechanics. Furthermore, very likely evaluable as true for *half* the particles measured to have negative spin component along axis one is the counterfactual "If the particle had not been measured at all, it would have been a particle with positive spin component along axis one." The truth of such a counterfactual together with the obvious regularity pattern constitutes the essence of what the author regards as a causal relationship, with measurement as cause and manifestation of negative spin component as effect.

Now let us *adjoin* the claim that particles do indeed carry (random) so-called "instruction" sets *at source emission*, and that the constitution of these instruction sets is influenced by the act of measurement. That is to say, we have a perfect example of what would happen if a cause were later than its effect, the cause being the quantum measurement and the effect being the (earlier) acquisition at source emission by the particle (and its correlated partner) of a quantum attribute value. It may be pleasing to some to note that acquisition of the same quantum attribute value by correlated particles can thereby be regarded as occurring when the particles are together at the source, rather than via instantaneous "information" transfer when they are distant. From this point of view a measurement does not simply cause a spin value to "be." It causes it to "have been." Philosophers generally agree that there is no inherent specification of time ordering in the notion of necessity of one event for another. Yet there has been no satisfactory resolution of what may be expected when the counterfactually necessary partner is the later one. Therefore, we have offered our own view.

Let us review briefly some consequences we have asserted would ensue from having a cause later than its effect. We have refuted the possibility of bilking reality by acting after an effect has occurred but before its cause has occurred so as to prevent the cause from happening. We have claimed that any bilking is done by reality to observers rather than vice versa, and that bilking involving acquisition of an altered past may be actual rather than counterfactual. Indeed, any time the (presumably randomly determined) values of the same quantum attribute for a correlated pair chance to constitute at source emission a state excluded from measured reality, an act of measurement introduces a reality change such that the excluded state never existed. (Given nonsimultaneous measurements at the same setting by detectors A and B of a correlated pair of particles, the causative measurement would be the earlier of the two measurements.)

One might metaphorically imagine a modest reality for which abundant evidence exists that it dresses unfashionably at times, yet that is never caught in unfashionable attire. As we stated in Section 2.2, we postulate a kind of bilking by reality that could not be detected by examination of memory or written records. In fact, it likely would be revealed to us only through seemingly anomalous results (involving the operation of reality change so-called "exclusion rules") from quantum mechanical experiments. (Apparently, one such exclusion rule could be stated "No reality reconstitution induced by an axis one measurement-as-R of a particle at a first detector A may include an axis one measurement-as-G of a correlated particle at a second detector B.")

We have seen that it is possible to talk consistently about changing the past if one makes the past a function of current epoch. Imagine emission of a particle pair at t_0, with a first quantum measurement of any kind on either member of the pair at t_f, and let t_i be such that $t_0<t_i<t_f$. Perhaps the quantum states ("instruction" sets) that appropriately may be assigned to the epoch t_0 *at the epoch* t_i for the particle pair are (randomly determined at t_0 from the perspective of t_i) RGR and GRG, respectively. But detection (measurement of spin along axis one) at t_f may alter reality in such a way that the quantum states that appropriately may be assigned to the epoch t_0 *at the epoch* t_f *for* the (now correlated) particle pair are RGG and RGR, respectively. (The quantum states are randomly determined at t_0 from the perspective of the reconstituted reality at t_f generated by a measurement which induces for the pair a common first element R.) The state-at-t_0 that should be attributed at t_i to an as-yet-unmeasured-at-t_i remote particle differs from the state-at-t_0 that should be attributed at t_f to that particle. "The past that is" at t_f is not at corresponding common epochs "the past that was" at t_i. What would be the ripple effects of such a successful, albeit inadvertent, action by experimenters to change the past?

The kind of reality change we postulate to be involved is not at all inconsistent. Intuition dictates that the laws of physics would apply throughout an appropriate span of spacetime in both the unreconstituted and the reconstituted realities associated with the occurrence of a cause later than its effect.

Perhaps an emotionally disturbing example, admittedly implausible even were backward causation possible at the measured, macroscopic level, will illustrate this last point. Suppose the foiling of an attempt to assassinate President Kennedy on November 22, 1963, is caused by the diversion of a certain meteor shower in the year 2020. Straining credulity a bit more, suppose the meteor shower diversion has no other effects in the past-as-we-know-it prior to that fateful day in 1963. In the ordinary course of events, we would find no reason nor means to divert the meteor shower, and the past-as-we-know-it would remain intact. However, a successful attempt to divert the meteor shower in the year 2020 would profoundly alter our collective past not only by reconstituting instantaneous casual chains at a certain epoch on November 22, 1963, but also via "rippling" through subsequent epochs. That is, the chains of forward causation that were unleashed in late 1963 would be replaced by more ordinary chains of forward causation such that, perhaps, John Kennedy would now be an aging professor emeritus at Harvard. (A "shorter" - and hence arguably more plausible - jump transition from the reality that we know in the 1990's would yield a reality involving President Kennedy's demise by some other means circa late 1963.)

More to the point, intuition (at least the author's) strongly dictates that among the laws of physics that are preserved across reality changes are *statistical* laws, e.g., a random run sequence that began with (RGR, GRG) would be reconstituted (via the backward causation associated with a first measurement in switch position one) as a *new* random sequence that began with, say, (RGG, RGR). It would not be reconstituted as an *otherwise* random sequence with first element constrained to be of the form (R??, R??).

In [3, pp. 128-134], we discussed the existence of an identifiable ordering of events that could be said to describe a kind of "sublocal" propagation internal to an occupied spacetime point. But it is not logically necessary that such propagation be so constrained. Propagation via a chain of counterfactual necessity could instantaneously yet sequentially (with respect to if, rather than with respect to time) transform an extended spacetime interval.

Consider that in the new history generated at t_f (as in the unreconstituted history that existed prior to t_f) it is a property of the past that *all* members of a sequence that was random in the unreconstituted reality were as yet unsampled at t_0. Therefore, it would be incorrect to claim that a key property of the first element is fixed (but as yet unbeknownst to anybody) at t_0, and that it is the data excluding the allegedly fixed property of the first element that ought to constitute a random array.

Instead, the replacement sequence would have the properties of a random array with respect to the *full* aggregation (for each measurement switch position). Production interior to the sequence of a particular 6-tuple of values (say (GRG, RGR)) associated with an ordered pair of "instruction" sets arguably would be governed by rules of conditional probability analogous to sequential selection from a bowl *without replacement.* A case can be made that successive measurements would reconstitute reality in such a way that the Bell inequality would be violated in just the quantitative manner predicted by quantum mechanics.

Let us return to the hypothetical and presumably incredible macroscopic bilking example we gave in which the foiling of an attempt on President Kennedy's life in 1963 is caused by diverting a meteor shower in 2020. Our purpose is to examine more closely our assumption that a successful attempt to preclude the meteor shower exacts its influence on 1963 via a kind of propagation. The question arises, Must the transformation of a spacetime interval be in accordance with the following?:

1. There is an instantaneous *chain* of counterfactual necessity acting backward through time and space, at once transforming the spacetime interval, but with successive links in the chain nonetheless necessarily spatiotemporally contiguous.

2. There is no spatial or temporal change in the location of the moving real-now arrow. (One does not "consume" the decades in question to regenerate them: They are at once "there.")

For example, consider the transformation according to statements 1 and 2 of a causal chain D(1), E(1), F(1) generated by spatiotemporal propagation with positive time flow. Suppose F(1) is followed by Z(1), where Z(1) is "backward causal" with respect to D(1). For Z(1) to exert its influence on D(1) via a chain of backward acting causes there must exist attributes $\tilde{E}$ and $\tilde{F}$ (in spatiotemporal coincidence with attributes E and F, respectively) so that we have the backward causal chain Z(1), $\tilde{F}(1)$, $\tilde{E}(1)$, D(1). Presumably, counterfactual deletion at the Z epoch of Z=1 would induce the real-now arrow to move vertically in the time-if plane, traversing a sequence of "trial" possible worlds. Each such candidate possible world would be examined (in some metaphysical sense) in turn before time was permitted to advance beyond the Z epoch. The candidate possible world sequence in question would be given by

(D, (E,$\widetilde{E}$), (F,$\widetilde{F}$), Z)= (1, (1,1), (1,1), 0); (1, (1,1), (1,0), 0); (1, (1,0), (1,0), 0); (0, (1,0), (1,0), 0); (0, (0,0), (1,0), 0); (0, (0,0), (0,0), 0).

Any absence of a chain of counterfactual necessity seems to involve instantaneous action at a spatial *and if* distance. To avoid a neo-Newtonian controversy over such a concept, perhaps a field theory based on space, time, and if should be devised. We suggest that a field theory defined over spacetimeif trajectories could have pragmatic value in some cases, e.g., modeling what "really" happens when a particle is annihilated from, or newly appears in, our reality. In the narrow context of EPR experiments, however, the issue of propagation can be sidestepped thanks to the high probability that (at least from the viewpoint of a quantum particle) source emission may properly be regarded as contiguous in space and time to detection.

The bilking introduction (or deletion) of a cause later than its effect may be regarded as introducing a discontinuity in the direction of the moving real-now arrow, which travels *vertically* in the time-if plane until it adopts an appropriate counterpart trajectory not forbidden by some exclusion rule.

The instantaneous transition introduces what we might term an "identity crisis" (discontinuity in sense of self) for the entity making the quantum jump parallel to the if axis. As we have noted, there would be (at best) sparce evidence available to such an entity that its so-called "actual" history is anything other than the bogus history that is actual for the counterpart that it has become.

Similarly, in the revisionist President Kennedy scenario, we would base our sense of self-identity upon a bogus past associated with a continuous path to the revised present, i.e., upon a past in accordance with our reconstituted written records and collective memory. In fact, in the general case, there would be philosophical grounds for rejecting as artificial the selection of any history other than the bogus one, except where the operation of an exclusion rule (as may happen in the EPR experiments) furnishes evidence that a jump transition has occurred.

In conclusion, we reiterate that an exclusion rule that seems to operate in Mermin's EPR experiment is one which imposes the following restriction upon the class of admissible worlds to which a particle can be transferred via imposition of a cause (quantum measurement of spin along axis one) later than its effect (quantum attribute acquisition at source emission): From a world containing a random sequence with first element of the form (R??, G??) transition must be made to a world containing a replacement random sequence with first element that accords with one of the two forms (R??, R??) or (G??, G??).

3.2 Nonlocality

Our nontraditional interpretation of how a backward acting cause would manifest itself introduces attributes each of which displays not just time-dependency of its truth value but, moreover, time-dependency of its truth value *at an epoch.* Accordingly, our supporting notation bears supplemental indexation. Thus, given a succession of instants t_0, t_1, t_2, ..., we denote by A_{ij} the truth value of attribute A at time t_i *as evaluated at time* t_j, where $j \geq i$, $i \geq 0$. Similarly, treating the world as an array of attributes distributed throughout space, we denote by W_{ij} the status of the world at time t_i *as evaluated at time* t_j, where $j \geq i$, $i \geq 0$. That is, for each epoch, we designate a *succession* of worlds (i.e., world descriptions) for that epoch. For example, at time t_1, we have the succession W_{11}, W_{12}, W_{13}, ...

The extreme possibilism inherent in our suggestion that inequalities like $A_{11} \neq A_{12}$ (and hence $W_{11} \neq W_{12}$) may be actualized at the quantum mechanical level, although it reopens in a grand way philosophical problems concerning transworld retention of self-identity, compensates mightily by helping to render intuitively evident the quantitative results of EPR experiments. It behooves us, therefore, to examine briefly some further consequences of this possibilism.

Suppose we treat the past prior to epoch t_1 as invariant (a possibility not excluded by our formalism), but that we actualize at time t_N (for some $N \geq 2$) a history for $t_1 \leq t < t_N$ that, in its $t_1 \leq t \leq t_i$ subsequence (for some $i < N$) was *counterfactual* at time t_i. It is conceptually advantageous to interpret this substitution (occurring instantly at t_N) of a new history for $t_1 \leq t < t_N$ as a history *regeneration* as would be associated with the *re-accessing* of epoch t_1 by a "phantom" moving real-now arrow.

Suppose further that we merge with this so-called "re-accessing" viewpoint the notion that each accessible version of the world has a nondeterministic component. It becomes sensible to question the assumption that a backward acting cause would manifest itself *only* via an earlier, spatially localized perturbation unleashing replacement causal chains, emanating in a forward time sense therefrom like eddy currents. Indeed, it seems feasible that a spatially localized perturbation at t_1 would be accompanied by an instantaneous global transformation at t_1 in the truth values of features that are nondeterministic in both the "original" world W_{11} and the current (at t_N) replacement world W_{1N}.

Now let us sketch a model with just enough formalism to highlight global behavior. Assume that the status of a world is a truth value array for an ordered set of attributes (distributed throughout space). Assume further that this set can be partitioned into an I-attribute subset (whose composition depends solely on the status of the world at the previous epoch, and whose member attributes have unconstrained, i.e., indeterministic, truth values) and an L-attribute subset (whose member attributes have truth values lawfully constrained to be interdependent, as well as dependent upon (1) contemporaneous I-values and (2) the status of the world at the previous epoch). Therefore, for $j \geq i \geq 1$, we may write

$$W_{ij} = (I_{ij} ; L_{ij})$$

To illustrate with a first example, we consider a world bearing just three spatially coincident attributes, namely the I-attribute J and the two L-attributes A and B. A traditional view for the successive epochs t_1 and t_2 is this: The two relevant world states are W_{11}= (J_{11};A_{11}, B_{11}) and W_{22}= (J_{22}; A_{22}, B_{22}), where functions f_A and f_B exist such that

$$A_{11} = f_A (B_{11}, J_{11}, W_{01})$$

$$B_{11} = f_B (A_{11}, J_{11}, W_{01})$$

$$A_{22} = f_A (B_{22}, J_{22}, W_{12})$$

$$B_{22} = f_B (A_{22}, J_{22}, W_{12})$$

Tacitly assumed along with $W_{01} = W_{00}$ is $W_{12} = W_{11}$, this latter comprising the set of assumptions $J_{12}=J_{11}$, $A_{12}=A_{11}$, $B_{12}=B_{11}$.

Our nontraditional view for the successive epochs t_1 and t_2 is this: The three relevant world states are $W_{11}=(J_{11};A_{11},B_{11})$, $W_{12}=(J_{12};A_{12},B_{12})$, and $W_{22}=(J_{22};A_{22},B_{22})$, where functions g_A and g_B exist such that

$$A_{12} = g_A\,(B_{12},J_{12},W_{02},W_{22})$$

$$B_{12} = g_B\,(A_{12},J_{12},W_{02},W_{22})$$

and where A_{11}, B_{11}, A_{22}, B_{22} are given by relations that agree, formally, with those used in the traditional approach. (Although the functions f_A and f_B match the corresponding f-functions of the traditional formulation, the behavior of their arguments is dissimilar. It should be emphasized that we allow $W_{11} \neq W_{12}$, perhaps involving all of $J_{11} \neq J_{12}$, $A_{11} \neq A_{12}$, $B_{11} \neq B_{12}$. On the other hand, the assumption of an invariant past prior to t_1 permits us to take $W_{00}=W_{01}=W_{02}$.)

Of course, for certain g-functions, there exists the theoretical possibility of loops or inconsistencies, e.g., for L-attribute values A_{12} and B_{12} dependent in a self-defeating way upon W_{02} and W_{22}. Yet this situation is not unprecedented: Inconsistencies could occur in the traditional formulation for certain f-functions, e.g., any for which $A_{11}=1 \varnothing B_{11}=1 \varnothing A_{11}=0$.

It concerns us at present to examine what is entailed by the possibility $I_{11} \neq I_{12}$. Toward that end, we consider as a second example a world having these features:

1. Indeterministic aspects are described by a three-vector of spatially distributed attributes.

2. There are two L-attributes. (This assures that the epoch t_1 can be re-accessed by the only means we have studied, namely backward causation.)

3. Epochs may be identified, for conceptual advantage, with increments of continuous time.

Specifically, suppose that we have two spatially coincident L-attributes A and B at location R0. Suppose further that at each location R1, R2, R3 there is a nucleus with a 50% chance of disintegrating at epoch t_1. The locations R0, R1, R2, R3 are mutually disjoint. Now imagine Joe, an omniscient observer (an artifice we can later happily discard) exercising an option to perform (1) nonperturbative global measurements at t_1, and (2) nonperturbative evaluations of the results of the measurements instantaneously at t_1, rather than retrodictively. According to Joe, it happens that decay occurs only at R2, i.e., with I-attributes defined in an obvious way by a J-vector, the appropriate J-vector is J_{11} = (0,1,0) or, in simpler notation, 010. At t_2, Joe exercises an option to function as an ordinary observer and imposes at R0 a backward cause B_{22}=1 (perhaps the act itself of receiving and processing ordinary signals from events at t_1!), yielding A_{12}=1 in place of A_{11}=0. (Reference to our first example should clarify that we have not postulated that backward causation is, of its essence, a product of finite time delay for signal acquisition. It is the retrospective assignment of simultaneous truth values for spatially separated attributes, in the current example, that leads us to consider finite time delay for signal acquisition.)

World W_{11} has been transformed into a world W_{12} in which, presumably, statistical laws continue to apply. Therefore, indeterminate processes in W_{12} are just that, and the re-accessing at t_2 of the t_1 epoch has opened the door to a (presumably randomly selected) value set J_{12} to replace J_{11}. Although the truth values A_{12} and B_{12} are fixed for world W_{12}, there are eight equally likely candidates for actualization as world W_{12}. These eight candidates correspond to eight equally likely value sets for J_{12}, namely 000, 001, 010, 011, 100, 101, 110, 111.

It is noteworthy that the multiplicity of possibilities for W_{12} reflects the admissibility of an instantaneous *global* transformation that accompanies the localized perturbation applied at R0 and involving attribute A. This virtual "action at a distance" involves instantaneous change of remote attribute truth values *from what they otherwise would have been*, not necessarily from what they were at the previous epoch. It involves instantaneous spatial "propagation" of local action: Intervention, applied at R0, also perturbs simultaneous outcomes at R1, R2, and R3. Denizens of world W_{12}, of course, do not directly experience dual realization of possibles (e.g., they will insist there was "always" a decay at R3 if J_{12}=001), though their indirect awareness of such phenomena may lead them to use counterfactuals in their language.

Presumably, the eight candidates for actualization as world W_{12} are all equally accessible at t_2 from the world W_{11} (partially characterized by J_{11}=010) that existed just prior to imposition of the backward acting cause. From a cosmic viewpoint, not one of these candidate worlds is "more neighboring" to the previously actualized world than is another.

Indeed, it is the assumption otherwise - that the outcome of remote indeterministic processes are biased toward similarity with their outcome had there been no local perturbation - that involves unwarranted localized control of remote phenomena. That local control of remote phenomena is unwarranted is illustrated by empirical quantum mechanical evidence that EPR methodology cannot be used for message transmittal. It is generally conceded that instantaneous *causal* spatial transmission is forbidden.

Our position, then, is that to regenerate a past epoch is to enter a world in which, with respect to a phantom moving now positioned at the past epoch, probabilistic events are not determined *anywhere*. Rather, in a spatial sense, it's a whole new ball game *everywhere*.

For enhanced intuitive attractiveness of this position, imagine for dialectical purposes that already measured phenomena can be undone, and consider this scenario: The author is effectuating that a replacement earlier version of himself invested in Texas Instruments as a fledgling company. Now surely he could not be doing so *secure in the knowledge that all relevant remote random economic factors contemporaneous with the investment action would remain constant,* so that the replacement version of himself thenceforth *of necessity* became wealthy.

To claim otherwise is to assert that the global retrospective afforded by time passage allows the author to know *precisely* what would be transpiring in Dallas as his younger counterpart in Houston took affirmative investment action. To claim otherwise is to assert that the author, strictly by a local disturbance, could transform his earlier self from a dynamic, unconstrained world into a counterpart in a largely predestined, nonlocally rigidly constrained replacement world. *If a surrogate world must "remember" and conform nonlocally to the incarnation it replaced, then the laws of physics (especially probabilistic laws) can hardly hold therein.*

3.3 Concluding remarks

It appears that an intuitively tractable model of nonlocality in quantum mechanics can be designed. Furthermore, the postulate that backward causation exists, *and manifests itself as analyzed in this treatise,* can also be applied to wave/particle duality and the uncertainty principle.

For example, we may associate with a single impact upon a phosphor screen an alteration of a particular electron's recent (unmeasured) past: At impact, the electron acquires the property that its trajectory from source to screen was "always" that of a particle. As our scheme involves time dependency of the location-at-an-epoch that should be attributed to a remote electron, this *complements,* rather than contradicts, the widely held view that the unmeasured electron travels from source to screen like a wave. (It does allow an essence, or self-identity, that bounces around arbitrarily and discontinuously rather than propagating ambiguously.) Indeed, incorporation of the widely held wave view into a backward causation model yields a wave/particle duality consistent with both the uncertainty principle and the Airy pattern of electron diffraction experiments. We simply reason thus: There being no alteration of the past associated with a nonimpact, no electron hits the screen in one of the dark rings because (in the unaltered past) the electron traveled from source to screen like a wave.

This backward-causation interpretation of the Airy pattern treats the acquisition of an altered past as rather more physical than bogus. It affords its new perspective on wave/particle duality by implicitly presupposing elements of a theory of knowledge which could be developed from the following metaphysical assumptions:

A1. *There is a world (called Real-World) that is objectively real.* Real-World consists of a "present" (actual) state plus an ordered sequence of "past" (actual) states, and is generated by an arrow of time that moves irreversibly and inexorably. "Acting to change the Real-World past" may optionally be regarded as an incoherent notion.

A2. *Real-World is highly chaotic.* "Jump discontinuities" in states are commonplace in Real-World, which may be likened to a succession of largely unrelated scenes. The laws of physics, as we know them, do not apply in Real-World.

A3. *Real-World is largely unknowable to observers.* Observers within Real-World function like funnels, or lenses, merging into sharp focus aspects of Real-World that are ordinarily diffuse and untidy. Self-identities are not preserved across Real-World jump discontinuities. It follows that a defining characteristic of observers is an "inability to see around corners." Accordingly, associated with each Real-World epoch (e.g., the present) is a phantom, monotonic history (called Observer-World), a product of rational interleaving of pooled observer data. (As the monotonic nature of Observer-World's history is achieved by brute force devices like shrinking meter sticks, slowing clocks, and ascribing magical properties to photons, perhaps "gerrymandering" is a more apt term than "interleaving.") Whereas Real-World's past is *objective*, Observer-World's past is *intersubjective.*

A4. *Observer-World is an aspect of a Real-World present.* A *record* for a current Real-World state consists of truth values for a spatial distribution of attributes. A record can be partitioned into two fields, namely (1) a *pure present field,* an array of truth values for what may be termed *vanguard attributes* , and (2) an *apparent history field,* an array of truth values for what may be termed *relic attributes.* Observer-World, at a Real-World epoch, is described by (1) its own pure present state for that epoch, together with (2) the sequential array of time-tagged data that defines the apparent history field for that epoch's record. Accordingly, Observer-World, though purportedly providing a tunnel-vision account of a Real-World epoch's *past*, more properly is regarded as an aspect of that Real-World epoch's *present.*

A5. *The "present" for an Observer-World associated with a current Real-World epoch must "agree with" the Real-World present.* For each Real-World epoch, the existence of an apparent history (i.e., a sense of how the current state "came to be") is intimately related to the existence of observers at that epoch. On the other hand, a sequence of past-to-present Real-World records, each stripped of its apparent history field, defines what may be termed *external reality.* Though, for a current Real-World epoch, the sequence of past-to-present states known as Observer-World may differ from the sequence of past-to-present states comprising external reality, Observer-World's present must *"match"* external reality's present.

A6. *Notwithstanding the highly chaotic nature of Real-World, there is causation therein.* Real-World responds to action by taking forms it would not have taken without action.

A7. *Ordinarily, external reality and Observer-World share a common recent evolution.* Action in Real-World generally can effectuate contemporaneous truth value changes only in spatially local vanguard attributes. Therefore, it is usual for external reality and Observer-World to "run along together," i.e., share the same recent past. Observer-World may be said to evolve, or "run along," when (as is ordinary) the Observer-World for a Real-World epoch is a "continuation" of the Observer-World for the previous Real-World epoch. (Incidentally, note that observers assign current truth values to spatially remote attributes *tentatively*, inferring them from relic attribute truth values.)

A8. *Certain Real-World actions, e.g., the making of quantum measurements, can effectuate truth value changes in relic attributes, not necessarily spatially local.* Any one such Real-World action can effectuate the acquisition of an Observer-World that differs in substantial historical ways from the Observer-World for the previous Real-World epoch. It is in this limited sense that action can "change the past." To say that the state-of-the-world-at-t_1 changed from t_2 to t_3, then, is to say that the state-at-t_1 of the Observer-World for Real-World epoch t_3 differed from the state-at-t_1 of the Observer-World for Real-World epoch t_2. (Interestingly, spatially remote attribute historical truth values are local in the sense that they are instantaneously available, i.e., known, to agents.)

A9. *The laws of physics apply in Observer-World.* Accordingly, an alternate name for Observer-World might be "the physical world." (There is room for distaste for a nomenclature that treats "real" and "metaphysical" as virtually synonymous.)

A10. *Self-identity determinations are made with respect to Observer-World.*

A11. *The laws of physics which hold in the Observer-World associated with a Real-World epoch ordinarily are the same laws of physics which held in the Observer-World associated with the previous Real-World epoch.* This, of course, is yet another assumption about Real-World, and hence is essentially unverifiable.

A12. *Knowledge, ordinarily, is with respect to Observer-World.* There is room for philosophical repugnance at (1) postulating the existence of an objective external reality, then (2) dignifying with the status of knowledge certain beliefs that do not accord with that external reality. But the alternative is this: There is only belief; there is virtually no knowledge. That alternative all-is-delusion doctrine has a certain nihilistic appeal to the author, who nevertheless leans toward the pragmatic practice of counting as bodies of knowledge coherent, consistent belief systems whose elements are verifiable, justified, and "true" with respect to a construct that contains the laws of physics. Here, "truth" is accordance with intersubjective records such as geological and astronomical data, memories, and written tomes.

Note that the indexation of knowledge does not involve us in counterfactuals. That is so, even given scenarios such as the following:

1. "At the previous epoch, A=1" is a true statement about the Real-World present (i.e., about the Observer-World associated with the Real-World present).

2. "At the previous epoch, A=0" is a true statement about the Real-World past, i.e., about external reality.

A13. *On rare occasions knowledge about external reality that "contradicts" knowledge about Observer-World becomes available to observers.* In general, acquisition at Real-World epoch t_2 of an Observer-World for which A=1 at earlier t_1 introduces self-identity discontinuities that preclude awareness that external reality was characterized by A=0 at Real-World epoch t_1. When A=0 at Real-World epoch t_1 was determined by *measurement* at that epoch, that measurement is effectively "undone" at t_2 for the observer, who (having no "recollection" that A=0 at t_1) experiences no confusion. Yet observer-experienced ambiguities may arise when knowledge of A's truth value at Real-World epoch t_1 was then acquired *other than by measurement.*

Now empiricism is replete with examples (half lives, kinetic theory of gases, Brownian motion) of knowing *about* the unknowable. Without paradox, then, we may assert that wave function collapse is a rare available factum about highly chaotic external reality. Surely it would be preposterous to ascribe to

Observer-World (which is governed by the laws of physics) the following happenstance:

1. Just prior to impact with a phosphor screen, an electron has an aspect about to strike a ring near the top of the screen, another aspect about to strike a ring near the bottom of the screen, and other aspects about to strike rings near the middle of the screen.

2. At the instant of screen impact, all those aspects make faster-than-light jumps to coalesce in the same unpredetermined tiny screen area.

As our unindexed statements generally refer to Observer-World, we *ought* to be saying, "The electron, prior to impact, was a particle." For consistency with the way we reason about macroscopic phenomena, we should reason thus: "The electron was a particle when it left the source. It was a particle when it hit the screen. Therefore, it was always a particle."

It is already conventional to treat the making of a quantum measurement as itself causal with respect to the acquisition of certain properties by quantum entities. We build upon this treatment as follows: Like most macroscopic events in Real-World, the impact of an electron with a phosphor screen has Real-World *effects.* That is, certain attributes are made to change their truth values with respect to the previous Real-World epoch. But among those affected attributes are relic attributes. Accordingly, the impact event "rectifies the past" by inducing the acquisition of an Observer-World different from external reality. In that Observer-World, the electron was always a particle.

But what about the fact that an aggregate of electrons that hit a screen produces a spread-out pattern of alternating light and dark concentric rings, even when the electrons are transmitted through a slit *one at a time*? Does not this evidence of diffraction and interference demonstrate that each electron disassociated into wavelets between source and screen?

In this regard, recall that we proposed the induction of an Observer-World different from external reality only for an *impact* event. We proposed no such thing for a *nonimpact.* Accordingly, dark rings exist because, in Real-World, electrons travel from source to screen like waves.

There is no electron that, except for a measurable self-interference event, would have hit the screen in one of the dark rings. Thus it remains legitimate to say, in the Observer-World we occupy after the pattern is complete, that undetected–and *only* undetected – electrons travel like waves. Yet it becomes legitimate to say, for that Observer-World, that a measured electron – when it was as yet unmeasured – traveled like a particle.

Symbolically, we may consider the attribute A = Electron wavelike? For a detected electron we have the following:

1. In external reality A traversed the truth value sequence (0, 1, 0).

2. In the Observer-World acquired at detection, A traversed the truth value sequence (0, 0, 0).

The upheaval in self-identity associated with the Observer-World acquisition is subtle enough to admit observer ambiguity regarding the appropriate intermediate truth value for A. At that critical intermediate epoch, in Real-World, there was an observer who correctly believed A=1. In his stead, at that critical epoch for the Observer-World acquired at the next Real-World epoch, there was an observer who incorrectly believed A=1.

There is no measurement that is "undone." It remains possible to claim that a measurement locks irreversibly into reality a quantum value at an epoch. It remains possible, *in that sense*, to assert that the past is immutable. But it becomes *impossible* to blend (1) infrequently acquired knowledge-about-external-reality coherently in the same pot with (2) more ordinary knowledge-about-Observer-World. It is just the attempt to do such blending that engenders observer confusion.

Numerous philosophical problems are introduced by our skeletal 13-item framework for a knowledge model of backward causation. A few matters deserving more rigorous analysis are these:

1. Does the indexation of knowledge "buy us anything" not available in more traditional formulations? (This matter was discussed briefly in Item A12.)

2. Must there be a unique mapping that picks out "like" epochs in the competing histories of external reality and Observer-World? (Would restricting attention to backward causation that left astronomical configurations unperturbed assure a common time scale for external reality and Observer-World?)

3. What can it mean to require that external reality and Observer-World "agree" at a current epoch? (Is there a shared vanguard attribute subrecord, such that the identity of respective truth value arrays assures this agreement? Must such a subrecord exclude every attribute, e.g., A= Emissary returning?, whose very definition presupposes a particular past? Are there *any* attributes whose very definition does not constrain the past?)

It is hoped that the reader who is impressed with the explanatory potential of this singular analysis of backward causation will undertake the investigation of such questions.

BIBLIOGRAPHY

[l] J. F. Allen, Towards a general theory of action and time, *Artificial Intelligence* 23 (1984) 123-154.

[2] I. Asimov, *The End of Eternity* (Random House, New York, NY, 1989).

[3] W. C. Bean, *Naive Causal Modeling* (Tito, Houston, TX, 2007).

[4] M. Black, Why cannot an effect precede its cause?, *Analysis* 16 (1956) 49-58.

[5] D. Bohm, *Causality and Chance in Modern Physics* (Van Nostrand, Princeton, NJ, 1957).

[6] M. Born, *Natural Philosophy of Cause and Chance* (Dover, New York, NY, 1964).

[7] R. Bradley and N. Swartz, *Possible Worlds* (Hackett, Indianapolis, IN, 1979).

[8] P. W. Bridgman, *The Logic of Modern Physics* (Ayer, Salem, NH, 1980).

[9] M. Bunge, *Causality and Modern Science* (Dover, New York, NY, 1979).

[10] R. Carnap, *Philosophical Foundations of Physics* (Basic Books, New York, NY, 1966).

[11] R. Carnap, *Meaning and Necessity: A Study in Semantics and Modal Logic* (University of Chicago Press, Chicago, IL, 1970).

[12] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence* (Addison-Wesley, Reading, MA, 1985).

[13] B. F. Chellas, *Modal Logic: an Introduction* (Cambridge University Press, New York, NY, 1980).

[14] R. M. Chisholm, The contrary-to-fact conditional, *Mind* 55 (1946) 289-307 .

[15] R. M. Chisholm and R. Taylor, Making things to have happened, *Analysis* 20 (1960) 73-78.

[16] R. G. Collingwood, Three senses of the word "cause", in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 118-126.

[17] A. Comte, *Introduction to Positive Philosophy*, trans. by F. Ferre (Hackett, Indianapolis, IN, 1988).

[18] J. de Kleer and J. S. Brown, A qualitative physics based on confluences, *Artificial Intelligence* 24 (1-3) (1984) 7-83 .

[19] J. de Kleer, How circuits work, *Artificial Intelligence* 24 (1-3) (1984) 205-280.

[20] B. d'Espagnat, The quantum theory and reality, *Scientific American* 241 (5) November (1979) 158-181.

[21] R. J. Doyle, Learning causal relations, in: T. M. Mitchell, J. G. Carbonell, and R. S. Michalski, eds., *Machine Learning: A Guide to Current Research* (Kluwer, Boston, MA, 1986) 55-58.

[22] C. J. Ducasse, *Nature, Mind and Death* (Open Court, La Salle, IL, 1951).

[23] C. J. Ducasse, Analysis of the causal relation, in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 142-153.

[24] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, *Physical Review* 47 (1935) 777-780.

[25] M. Dummett, Can an effect precede its cause?, in: *Truth and Other Enigmas* (Harvard University Press, Cambridge, MA, 1978) 319-332.

[26] J. H. Fetzer, *Scientific Knowledge: Causation, Explanation and Corroboration* (Kluwer Academic, Norwell, MA, 1981).

[27] A. Flew, Causal disorder again, *Analysis* 17 (1957) 81-86.

[28] D. Follesdal, Quantification into causal contexts, *Boston Studies in the Philosophy of Science* 2 (1965) 263-274.

[29] K. D. Forbus, Qualitative process theory, *Artificial Intelligence* 24 (1-3) (1984) 85-168.

[30] D. Gasking, Causation and recipes, in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 126-132.

[31] M. L. Ginsberg, Counterfactuals*, Artificial Intelligence* 30 (1986) 35-79.

[32] M. L. Ginsberg and D. E. Smith, Reasoning about action I: a possible worlds approach, *Artificial Intelligence* 35 (1988) 165-195.

[33] K. Gödel, A remark about the relationship between relativity theory and idealistic philosophy, in: P. A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist* (Open Court, La Salle, IL, 1988) 557-562.

[34] N. Goodman, The problem of counterfactual conditionals, *Journal of Philosophy* 44 (1947) 113-128.

[35] H. L. A. Hart and A. M. Honoré, *Causation and the Law* (Oxford University Press, New York, NY, 1959).

[36] P. J. Hayes, The naive physics manifesto, in: D. Michie, ed., *Expert Systems in the Microelectronic Age* (Edinburgh University Press, Edinburgh, 1979).

[37] C. G. Hempel, *Philosophy of Natural Science* (Prentice-Hall, Englewood Cliffs, NJ, 1966).

[38] I. Hinckfuss, *The Existence of Space and Time* (Clarendon Press, Oxford, 1975).

[39] T. Hobbes, *Nature and Mind: Selected Essays* (Russell and Russell, New York, NY, 1965).

[40] P. Horwich, *Asymmetries in Time* (MIT Press, Cambridge, MA, 1989).

[41] D. Hume, *An Enquiry Concerning Human Understanding* (Open Court, Peru, IL, 1988).

[42] E. H. Hutten, *The Language of Modern Physics* (Macmillan, New York, NY, 1956).

[43] P. Jackson, On the semantics of counterfactuals, in: *Proceedings 11th International Joint Conference on Artificial Intelligence*, Detroit, MI, 2 (1989) 1382-1387.

[44] A. C. Kak, Spatial reasoning, *AI Magazine* 9 (2) June (1988) 23.

[45] N. Kemp-Smith, *Commentary to Kant's Critique of Pure Reason* (Humanities, Atlantic Highlands, NJ, 1962).

[46] W. Kneale, Natural laws and contrary-to-fact conditionals, in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 46-49.

[47] W. Kneale, Universality and necessity, in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 53-63.

[48] B. Kuipers, Commonsense reasoning about causality: deriving behavior from structure, *Artificial Intelligence* 24 (1-3) (1984) 169-203.

[49] D. Lewis, Counterpart theory and quantified modal logic, *Journal of Philosophy* 65 (5) (1968) 113-136.

[50] D. Lewis, *Counterfactuals* (Harvard University Press, Cambridge, MA, 1973).

[51] D. Lewis, Causation, *Journal of Philosophy* 70 (1973) 556-567.

[52] D. Lewis, Counterfactual dependence and time's arrow, *Nous* 13 (1979) 455-476.

[53] M. J. Loux, ed., *The Possible and the Actual: Readings in the Metaphysics of Modality* (Cornell University Press, Ithaca, NY, 1988).

[54] D. McDermott, A temporal logic for reasoning about processes and plans, *Cognitive Science* 6 (1982) 101-155.

[55] E. Mach, *The Science of Mechanics,* trans. by T. J. McCormack (Open Court, La Salle, IL, 1942).

[56] J. L. Mackie, Counterfactuals and causal laws, in: T. L. Beauchamp, ed., *Philosophical Problems of Causation* (Dickenson, Encino, CA, 1974) 95-105.

[57] J. L. Mackie, *The Cement of the Universe: A Study Of Causation* (Oxford University Press, New York, NY, 1980).

[58] J. M. E. McTaggert, *Philosophical Studies* (Ayer, Salem, NH, 1934).

[59] Z. Manna and A. Pnueli, Verification of concurrent programs: the temporal framework, in: R. S. Boyer and J. S. Moore, eds., *The Correctness Problem in Computer Science* (Academic Press, San Diego, CA, 1981).

[60] H. Mehlberg, *Time, Causality, and the Quantum Theory*, 2 volumes (Dordrecht: Reidel, Boston, MA, 1980).

[61] N. D. Mermin, Is the moon there when nobody looks? Reality and the quantum theory, *Physics Today* 38 (4) April (1985) 38-47 .

[62] E. Meyerson, *Identité et Réalité* (Alcan, Paris, 1908).

[63] J. S. Mill, *A System of Logic: Ratiocinative and Inductive*, 2 volumes, J. M. Robson, ed. (University of Toronto Press, Cheektowaga, NY, 1974).

[64] E. Nagel, *The Structure of Science* (Hackett, Indianapolis, IN, 1979).

[65] D. Pears, The priority of causes, *Analysis* 17 (1957) 54-63.

[66] K. Pearson, *The Grammar of Science* (Adam and Charles Black, London, 1911).

[67] H. Poincaré, *Science and Hypothesis* (Dover, New York, NY, 1905).

[68] W. V. Quine, *Word and Object* (MIT Press, Cambridge, MA, 1960).

[69] W. V. Quine, *From a Logical Point of View* (Harvard University Press, Cambridge, MA, 1961).

[70] H. R. Reichenbach and M. Reichenbach, *The Direction of Time* (University of California Press, Berkeley, CA, 1971).

[71] T. Rothman, The seven arrows of time, *Discover* 8 (2) February (1987) 63-77.

[72] B. Russell, On the notion of cause, in: *Mysticism and Logic* (Doubleday, Garden City, NY, 1951) 174-201.

[73] B. Russell, *Our Knowledge of the External World* (Humanities Press, Atlantic Highlands, NJ, 1972).

[74] W. C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton University Press, Princeton, NJ, 1984).

[75] M. Schlick, Causality in everyday life and in recent science, in: H. Feigl and W. Sellars, eds., *Readings in Philosophical Analysis* (Ridgeview, Atascadero, CA, 1981).

[76] A. Schopenhauer, *The Fourfold Root of the Principle of Sufficient Reason*, trans. by E. F. J. Payne (Open Court, Peru, IL, 1974).

[77] M. Scriven, Randomness and the causal order, *Analysis* 17 (1957) 5-9.

[78] C. Selltiz, M. Jahoda, M. Deutsch, and S. W. Cook, Studies testing causal hypotheses, Chapter 4 in: *Research Methods in Social Relations* (Holt, Rinehart, and Winston, Chicago, IL, 1959) 79-143.

[79] Y. Shoham, Reasoning about causation in knowledge-based systems, in: *Proceedings 2nd Conference on Artificial Intelligence Applications: The Engineering of Knowledge-Based Systems*, Miami

Beach, FL (IEEE Comput. Soc. Press, Washington, DC, 1985) 629-634.

[80] Y. Shoham, *Reasoning About Change* (MIT Press, Cambridge, MA, 1988).

[81] Y. Shoham and N. Goyal, Temporal reasoning in artificial intelligence, in: H. Shrobe, ed., *Frontiers of Artificial Intelligence* (Morgan-Kaufmann, San Mateo, CA, 1988).

[82] Y. Shoham, Time for action: on the relation between time, knowledge, and action, in: *Proceedings 11th International Joint Conference on Artificial Intelligence*, Detroit, MI, 2 (1989) 954-959, 1173 .

[83] H. A. Simon, Causal ordering and identifiability, in: D. Lerner, ed., *Cause and Effect* (Collier-Macmillan, Toronto, 1965).

[84] L. Sklar, *Space, Time, and Spacetime* (University of California Press, Berkeley, CA, 1974).

[85] J. L. C. Smart, *Between Science and Philosophy* (Random House, New York, NY, 1968).

[86] R. Stalnaker, A theory of conditionals, in: N. Rescher, ed., *Studies in Logical Theory* (Oxford University Press, New York, NY, 1968).

[87] J. Sticklen and B. Chandrasekaran, Integrating Classification-Based Compiled Level Reasoning With Function-Based Deep Level Reasoning, in: W. Horn, ed., *Causal AI Models: Steps Toward Applications* (Hemisphere, Washington, DC, 1990) 191-220.

[88] Y. P. Terletskii, *Paradoxes in the Theory of Relativity* (Plenum Press, New York, NY, 1968).

[89] R. Tucker, *Logics for Artificial Intelligence* (John Wiley, New York, NY, 1984).

[90] B. C. van Fraassen, *The Scientific Image* (Clarendon Press, Oxford, 1980).

[91] G. H. von Wright, *Causality and Determinism* (Columbia University Press, New York, NY, 1974).

[92] G. J. Whitrow, *The Natural Philosophy of Time* (Oxford University Press, New York, NY, 1980).

[93] L. Wittgenstein, *Tractatus Logico-Philosophicus*, trans. by D. F. Pears and B. F. McGuinness (Routledge & Kegan Paul, London, 1966).

[94] P. J. Zwart, *About Time: A Philosophical Inquiry Into the Origin and Nature of Time* (North Holland, Amsterdam, 1976).

INDEX