TECHNICAL REPORT
R-69
June 1988

Research Note

Embracing Causality in Default Reasoning*

Judea Pearl
Cognitive Systems Laboratory, UCLA Computer Science Department, Los Angeles, CA 90024, U.S.A.

Recommended by E. Davis and C. Habel

ABSTRACT

The purpose of this note is to draw attention to certain aspects of causal reasoning which are pervasive in ordinary discourse yet, based on the author's scan of the literature, have not received due treatment by logical formalisms of common-sense reasoning. In a nutshell, it appears that almost every default rule falls into one of two categories: expectation-evoking or explanation-evoking. The former describes association among events in the outside world (e.g., fire is typically accompanied by smoke); the latter describes how we reason about the world (e.g., smoke normally suggests fire). This distinction is consistently recognized by people and serves as a tool for controlling the invocation of new default rules. This note questions the ability of formal systems to reflect common-sense inferences without acknowledging such distinction and outlines a way in which the flow of causation can be summoned within the formal framework of default logic.

1. How Old Beliefs Were Established Determines Which New Beliefs Are Evoked

Let A and B stand for the following propositions:

A "Joe is over 7 years old."
B "Joe can read and write."

Case 1. Consider a reasoning system with the default rule

$$\text{def}_B : B \rightarrow A .$$

* This is a revised version of a paper presented at the AAAI-87 Conference, Seattle, WA.

A new fact now becomes available,

e_1 "Joe can recite passages from Shakespeare,"

together with a new default rule:

$$\text{def}_1 : e_1 \rightarrow B .$$

Case 2. Consider a reasoning system with the same default rule,

$$\text{def}_B : B \rightarrow A .$$

A new fact now becomes available,

e_2 "Joe's father is a Professor of English,"

together with a new default rule,

$$\text{def}_2 : e_2 \rightarrow B .$$

(To make def_2 more plausible, one might add that Joe is known to be over 6 years old and is not a moron.)

Common sense dictates that Case 1 should lead to conclusions opposite to those of Case 2. Learning that Joe can recite Shakespeare should evoke belief in Joe's reading ability, B, and, consequently, a correspondingly mature age, A. Learning of his father's profession, on the other hand, while still inspiring belief in Joe's reading ability, should *not* trigger the default rule $B \rightarrow A$ because it does not support the hypothesis that Joe is over 7. On the contrary;

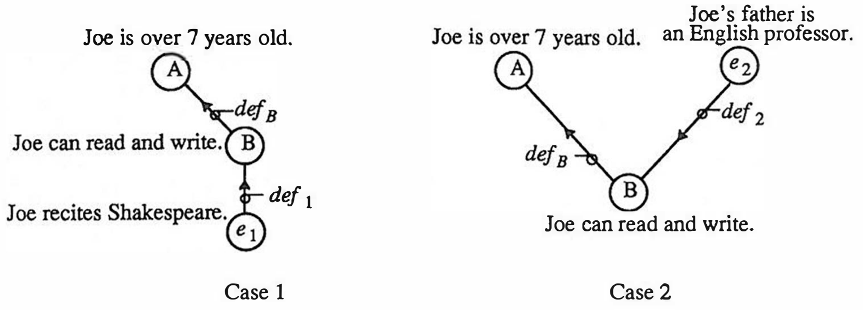

FIG. 1. The default rule $B \rightarrow A$ should be invoked when B is established by evidential information (Case 1) and inhibited when B is established by prediction (Case 2). Causal rules point downwards and evidential rules upwards.

whatever evidence we had of Joe's literary skills could now be partially attributed to the specialty of his father rather than to Joe's natural state of development. Thus, if a belief were previously committed to A, and if measures of belief were permitted, it would not seem unreasonable that e_2 would somewhat *weaken* the belief in A.

From a purely syntactic viewpoint, Case 1 is identical to Case 2. In both cases we have a new fact triggering B by default. Yet, in Case 1 we wish to encourage the invocation of $B \rightarrow A$ while, in Case 2, we wish to inhibit it. Can a default-based reasoning system distinguish between the two cases?

The advocates of existing systems may argue that the proper way of inhibiting A in Case 2 would be to employ a more elaborate default rule, where more exceptions are stated explicitly. For example, rather than $B \rightarrow A$, the proper default rule should read: $B \rightarrow A \mid \text{UNLESS } e_2$. Such exceptions can be encoded as nonnormal defaults in Reiter's logic [11] or as *out* justifiers in truth-maintenance systems [2].

Unfortunately, this cure is inadequate on several grounds. First, it requires that every default rule be burdened with an unmanageably large number of conceivable exceptions. Second, it misses the intent of the default rule $\text{def}_B : B \rightarrow A$, the primary aim of which was to evoke belief in A whenever the truth of B can be ascertained; it would be very disturbing for the rule author having to dream up far-fetched exceptions instead of simply articulating everyday knowledge that children with reading ability are typically over seven years old. Third, while correctly inhibiting A in Case 2, the UNLESS cure would also inhibit A in many other cases where it should be encouraged. For example, suppose we actually test Joe's reading ability and find out that it is at the level of a 10-year old child, unequivocally establishing the truth of B. Are we to suppress the natural conclusion that Joe is over 7 on the basis of his father being an English professor? There are many other conditions under which even a 5-year-old boy can be expected to acquire reading abilities, yet, these should not be treated as exceptions in the default-logical sense because those same conductive conditions are also available to a 7-year old; and, consequently, they ought not to preclude the natural conclusion that a child with reading ability is, typically, over 7. They may lower, somewhat, our confidence in the conclusion but should not be allowed to totally and permanently suppress it.

To summarize, what we want is a mechanism that is sensitive to how B was established. If B is established by direct observation or strong evidence supporting it (Case 1), the default rule $B \rightarrow A$ should be invoked. If, on the other hand, B was established by *expectation*, *anticipation* or *prediction* (Case 2), then $B \rightarrow A$ should not be invoked, no matter how strong the expectation.

The asymmetry between expectation-evoking and explanation-evoking rules is not merely that of temporal ordering, but is more a product of human memory organization. For example, age evokes expectations of certain abilities not because it precedes them in time (in many cases it does not) but because the concept called "child of age 7" was chosen by the culture to warrant a name for bona fide frame, while those abilities were chosen as expectational slots in that frame. Similar asymmetries can be found in object-property, class-subclass and action-consequence relationships.

2. More on the Distinction between Causal versus Evidential Support

Consider the following two sentences:

> Joe seemed unable to stand up; so, I believed he was injured. (1)
>
> Harry seemed injured; so, I believed he would be unable to stand up. (2)

Any reasoning system that does not take into account the direction of causality or, at least, the source and mode by which beliefs are established is bound to conclude that Harry is as likely to be drunk as Joe. Our intuition, however, dictates that Joe is more likely to be drunk than Harry because Harry's inability to stand up, the only indication for drunkenness mentioned in his case, is portrayed as an expectation-based property emanating from injury, and injury is a perfectly acceptable alternative to drunkenness. In Joe's case, on the other hand, not-standing-up is described as a primary property supported by direct observations, while injury is brought up as an explanatory property, inferred by default.

Note that the difference between Joe and Harry is not attributed to a difference in our confidence in their abilities to stand up. Harry will still appear less likely to be drunk than Joe when we rephrase the sentences to read:

> Joe showed slight difficulties standing up; so, I believed he was injured. (1′)
>
> Harry seemed injured; so, I was sure he would be unable to stand up. (2′)

Notice the important role played by the word "so." It clearly designates the preceding proposition as the primary source of belief in the proposition that follows. Natural languages contain many connectives for indicating how conclusions are reached (e.g., therefore, thus, on the other hand, nevertheless, etc.). Classical logic, as well as known versions of default logic, appears to stubbornly ignore this vital information by treating all believed facts and facts derived from other believed facts on equal footing. Whether beliefs are established by external means (e.g., noisy observations), by presumptuous expectations, or by quest for explanation does not matter.

But even if we are convinced of the importance of the sources of one's belief; the question remains how to store and use such information. In the Bayesian analysis of belief networks [5], this is accomplished using numerical parameters; each proposition is assigned two parameters, π and λ, one measuring its accrued *causal* support and the other its accrued *evidential* support. These parameters then play decisive roles in routing the impacts of new evidence throughout the network. For example, Harry's inability to stand up will accrue some causal support, emanating from injury, and zero evidential support, while Joe's story will entail the opposite support profile. As a result, having observed blood stains on the floor would contribute to a reduction in the overall belief that Joe is drunk but would not have any impact on the belief that Harry is drunk. Similarly, having found a whiskey bottle nearby would weaken the belief in Joe's injury but leave no impact on Harry's.

These inferences are in harmony with intuition. Harry's inability to stand up was a purely conjectural expectation based on his perceived injury, but it is unsupported by a confirmation of any of its own, distinct predictions. As such, it ought not to pass information between the frame of injury and the frame of drunkenness. The mental act of imagining the likely consequences of an hypothesis does not activate other, remotely related, hypotheses just because the latter could also cause the imagined consequence. For an extreme example, we would not interject the possibility of a lung cancer in the context of a car accident just because the two (accidents and cancer) could lead to the same eventual consequence—death.

Can a nonnumeric logic capture and exploit these nuances? I think, to some degree, it can. True, it cannot accommodate the notions of "weak" and "strong" expectations, nor the notion of "accrued" support, but this limitation may not be too severe in some applications, e.g., one in which belief or disbelief in a proposition is triggered by just a few decisive justifications. What we can still maintain, though, is an indication of how a given belief was established—by expectational or evidential considerations, or both, and use these indications for deciding which default rules can be activated in any given state of knowledge.

3. The *C-E* System: A Coarse Logical Abstraction of Causal Directionality

Evidently, common-sense reasoning involves two types of default rules: expectation evoking (e.g., if fire then smoke), and explanation evoking (e.g., if smoke then fire). We call the first *causal* rules, and the second *evidential* rules. The semantics that default logics associate with a default rule $A \rightarrow B$ is usually given in terms of a license to presume B, if A is believed, as long as B is consistent with currently held beliefs [11]. This semantics fits the nature of causal rules but not that of evidential rules. Rules of the type, "if observation A then hypothesis B," should be blocked not only when B is inconsistent with current beliefs, but also whenever an alternative explanation is available for the observation A, even when B is perfectly consistent with current beliefs. The problems described in the preceding sections stem from subjecting the two types of defaults to the same operational semantics: presume B unless it is contradictory.

There are two ways of handling these problems; one is to admit default rules of only one kind, the second is to admit a mixture of causal and evidential rules, tab each rule by its type, and manage them accordingly. The first method is certainly easier to implement. The MYCIN [13] system, for example, admits only evidential rules (always pointing from evidence to hypothesis); it can perform simple diagnoses but cannot combine diagnosis with prediction [12]. Alternatively, one can admit as input only causal rules, as is indeed the prevailing practice in Bayes' analysis; input information is given in a if-cause-then-effect format, while diagnoses are *derived* by explanation-seeking procedures (e.g. minimization), rather than by explicit diagnostic rules [6, 7]. Poole [10] has, likewise, devised a logic-based system where default rules are restricted to causal type, and reasoning from evidence to hypotheses is accomplished by specialized "theory formation" procedures. Such causal-based systems (often called "model-based" or "first-principles-based") enjoy the features of parsimony, stability and modularity, and facilitate a more natural, declarative representation of world knowledge.

In practice, however, most default-handling systems in AI admit a mixture of causal and evidential rules. For example, truth-maintenance systems would accept both causal justifiers, as in (IS FRED PROFESSOR) → (POOR FRED), and evidential justifiers as in (PAIN FRED SIDE) → (HAS FRED APPENDICITIS) [1]. The reason being that, despite the advantages of causal systems, it is hard for rule authors to resist the temptation of articulating compiled procedural knowledge, leading from familiar situations to previously successful actions or guesses, e.g., that smoke suggests fire, that symptons suggest diseases, etc. The C-E logic proposed here is an attempt to maintain plausibility in a mixed system, where causal and evidential rules reside side by side, each labeled by its type.

Let each default rule in the system be labeled as either C-def (connoting "causal") or E-def (connoting "evidential"). The former will be distinguished by the symbol $\rightarrow_C$, as in "FIRE$\rightarrow_C$SMOKE," meaning "FIRE causes SMOKE," and the latter by $\rightarrow_E$, as in "SMOKE $\rightarrow_E$FIRE," meaning "SMOKE is evidence for FIRE." Correspondingly, let each believed proposition be labeled by a distinguishing symbol, "E" *or* "C." A proposition P is E-believed, written $E(P)$, if it is a direct consequence of some E-def rule. Otherwise, if P can be established as a direct consequence of only C-def rules, it is said to be C-believed, written $C(P)$, supported solely by expectation or anticipation. The semantics of the C-E distinction are captured by the following three inference rules:

$$\text{(a)}\quad \frac{\begin{array}{r}P \rightarrow_C Q\\ C(P)\end{array}}{C(Q)} \qquad \text{(b)}\quad \frac{\begin{array}{r}P \rightarrow_C Q\\ E(P)\end{array}}{C(Q)} \qquad \text{(c)}\quad \frac{\begin{array}{r}P \rightarrow_E Q\\ E(P)\end{array}}{E(Q)}$$

Note that we purposely precluded the inference rule:

$$\frac{\begin{array}{r}P \rightarrow_E Q\\ C(P)\end{array}}{Q}$$

which led to counter-intuitive conclusions in Case 2 of Joe's story.

Inference rules (a), (b) and (c) imply that E-believed conclusions can only attain E-believed status by a chain of purely E-def rules. C-believed conclusions, on the other hand, may be obtained from a mix of C-def and E-def rules. For example, an E-def rule may (viz., (c)) yield an E-believed conclusion which can feed into a C-def rule (viz., (b)) and yield a C-believed conclusion. Note, also, that the three inference rules above would license the use of $A \rightarrow B$ and $B \rightarrow A$ without falling into the circular reasoning trap. Iterative application of these two rules would never cause a C-believed proposition to become E-believed because at least one of the rules must be of type C.

The distinction between the two types of rules can be demonstrated using the following example (see Fig. 2).

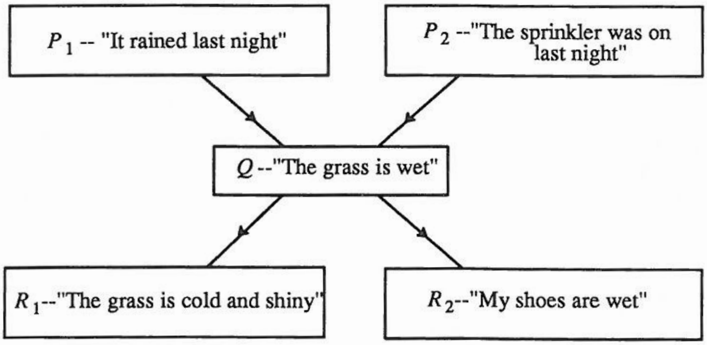
P_1 -- "It rained last night"
P_2 --"The sprinkler was on last night"
Q --"The grass is wet"
R_1--"The grass is cold and shiny"
R_2--"My shoes are wet"

FIG. 2. P_1 is invoked as an explanation when Q is established by observing R_1 or R_2, but not by observing P_2.

Let P_1, P_2, Q, R_1, and R_2 stand for the propositions:

P_1	"It rained last night,"
P_2	"The sprinkler was on last night,"
Q	"The grass is wet,"
R_1	"The grass is cold and shiny,"
R_2	"My shoes are wet."

The causal and evidential relationships between these propositions would be written:

$$\begin{array}{ll} P_1 \rightarrow_C Q\,, & Q \rightarrow_E P_1\,, \\ P_2 \rightarrow_C Q\,, & Q \rightarrow_E P_2\,, \\ Q \rightarrow_C R_1\,, & R_1 \rightarrow_E Q\,, \\ Q \rightarrow_C R_2\,, & R_2 \rightarrow_E Q\,. \end{array}$$

If Q is established by an E-def rule such as $R_1 \rightarrow_E Q$, then it can trigger both P_1 as explanation, and R_2 as prediction. However, if Q is established merely by a C-def rule, say $P_2 \rightarrow_C Q$, then it can trigger R_2 (and R_1) but *not* P_1.

The essence of the causal asymmetry stems from the fact that two causes of a common consequence interact differently than two consequences of a common cause; in the absence of direct links between the two, the former *compete* with each other, while the latter *support* each other. Moreover, the former interact when their connecting proposition is *confirmed*, the latter interact only when their connecting proposition is *unconfirmed*. In our example, the state of the sprinkler would influence our belief in rain only when the grass wetness is confirmed by observation. However, knowing whether the shoes are wet or dry can influence the prediction "the grass is cold" only prior to confirming the wetness of the grass. A logic of causal dependencies is given in [9].

Let us see how this C-E system resolves the problem of Joe's age (see Fig. 1). def_B and def_1 will be classified as E-def rules, while def_2 will be proclaimed a C-def rule. All provided facts (e.g., e_1 and e_2) will naturally be E-believed. In Case 1, B will become E-believed (via rule (c)) and, subsequently, after invoking def_B in rule (c), A, too, will become E-believed. In Case 2, however B will only become C-believed (via rule (b)) and, as such, cannot invoke def_B, leaving A undetermined, as expected.

To handle retraction we can employ a mechanism of "justification maintenance," similar to that used in truth-maintenance systems [2]. We define an *extension* to be an assignment of $C/E/$OUT status to the propositions in the system that is closed under rules (a), (b) and (c). An extension X is said to be *well-founded* if all its labels could be justified by the three inference rules above. In other words, every E-believed proposition Q in X is either given as a fact or is a conclusion of some E-def rule $P \rightarrow_E Q$ where P is labeled E; every C-believed proposition Q in X is a conclusion of some C-def rule $P \rightarrow_C Q$ where P is labeled either E or C. Newly added facts propagate their impact on the beliefs of other propositions by maintaining the well-foundedness of the extension. For example, if in Joe's story we first learn facts e_1 and e_2, then the only well-founded extension is $X_1 = \{E(e_1), E(e_2), E(B), E(A)\}$, namely all propositions are E-believed. If we later learn a new fact that suppresses the default def_2 (e.g., e_3, "Joe is blind and always repeats what he hears"), then extension X_1 gives place to $X_2 = \{E(e_1), E(e_2), E(e_3), C(B), \text{OUT}(A)\}$. Thus, suppressing def_1 causes B to become C-believed (causally justified by the truth of e_2) which further suppresses the rule def_B, and retracts the belied in A—Joe's being over 7 years old.

The merits of this definition of well-foundedness can be demonstrated when applied to the so-called "Yale shooting problem" [3]. In its simplest version, the problem involves shooting a person known to be alive at time t_1 ($\text{ALIVE}(t_1)$) with a gun known to be loaded at t_0 ($\text{LOADED}(t_0)$). Normally, we would expect the gun to remain loaded at t_1 and the victim to be dead at time t_2. Yet, if one expresses the natural tendency of things to persist over time by the default rules:

$$\text{LOADED}(t_0) \rightarrow \text{LOADED}(t_1)\,, \tag{3}$$

$$\text{ALIVE}(t_1) \rightarrow \text{ALIVE}(t_2)\,, \tag{4}$$

and the impact of shooting a loaded gun, by the rule:

$$\text{LOADED}(t_1) \rightarrow \text{DEAD}(t_2) \tag{5}$$

an anomalous extension ensues, whereby the victim is alive at t_2 and the gun is unloaded at t_1. The anomalous extension is assembled by applying rule (4) to the fact $\text{ALIVE}(t_1)$, followed by the contrapositive form of rule (5)

$$\text{ALIVE}(t_2) \rightarrow \neg \text{LOADED}(t_1)\,. \tag{5'}$$

The way the C-E system handles this problem would be to label (3)–(5) as causal rules and (5′) as an evidential rule. Starting with the facts $\text{LOADED}(t_0)$ and $\text{ALIVE}(t_1)$, if we first apply rules (3) and (5), we get the intended well-founded extension

$$X_1 = \{E[\text{LOADED}(t_0)], E[\text{ALIVE}(t_1)]\,, C[\text{LOADED}(t_1)], C[\text{DEAD}(t_2)]\}\,,$$

while applying rules (3) and (4) yields another anomalous well-founded extension:

$$X_2 = \{E[\text{LOADED}(t_0)], E[\text{ALIVE}(t_1)], C[\text{LOADED}(t_1)], C[\text{ALIVE}(t_2)]\} .$$

The anomalous extension of Hanks and McDermott, entaining $\neg\text{LOADED}(t_1)$, would not be well-founded because $\text{ALIVE}(t_2)$ is C-believed, hence, it cannot serve as a justification for an E-def rule like (5′).

Although the C-E system yields an anomalous extension X_2, it is not an unreasonable extension considering the syntax of the rules used. Indeed, exchanging $\text{ALIVE}(t_1)$ with the predicate $\text{WEARING-BULLETPROOF-VEST}(t_1)$ would satisfy rules (3)–(5) and (5′) and would render X_2 a more acceptable extension than X_1. In other words, there is no syntactic way of inferring from rules (3)–(5) that being alive at t_1 does not constitute protection against gun fire. (The purpose of rule (5) would then be to assert that dead people cannot be revived by being shot). To convey the disruptive effect of gun fire over the persistence of life, one can use, for example, the rule

$$\text{ALIVE}(t_1) \wedge \text{LOADED}(t_1) \rightarrow \text{DEAD}(t_2)$$

instead of (5) (see [7]). However, unlike the reasoning presented by Hanks and McDermott and regardless of the mechanism one chooses to represent the volatility of life under gun shots, the C-E logic will never allow the hypothetical prediction $\text{ALIVE}(t_1) \rightarrow \text{ALIVE}(t_2)$ to turn backwards and trigger doubts in the loadness of the gun at t_1. Again, this asymmetry is not unique to temporal ordering but is applicable to property inheritance and class-subclass relationships in general [4].

4. Implicit Suppressors and the Need for Finite Abstractions

E-believed status enjoys some advantages over C-believed status. The former can invoke both C-def and E-def rules, while the latter, no matter how strong the belief, invokes only C-def rules. On the other hand, C-def rules are more powerful than E-def rules, since the former can be applied to both E-believed and C-believed propositions, while E-def rules can be applied only to E-believed propositions. More generally, E-def rules are weaker because they can be undermined by propositions that, in themselves, do not contradict nor oppose the conclusion of the rules, if only they offer alternative explanations for the antecedent. In Fig. 2, for example, P_2 deactivates the rule $Q \rightarrow P_1$ despite the fact that is is perfectly consistent for a sprinkler, P_2, to turn on on a rainy night, P_1. This suppression reflects the natural tendency of people to prefer simpler explanations (i.e., involving fewer assumptions), and, hence, can be regarded as a local filtering scheme, serving some grand minimization policy.

The computational advantages of such suppression can be demonstrated in the context of the "frame problem" associated with the E-def rule: "If the car does not start, assume the battery is dead." Obviously, there are many exceptions to this rule, e.g., ". . . unless the starter is burned," ". . . unless someone pulled the spark plugs," ". . . unless the gas tank is empty," etc., and, if any of these conditions is believed to be true, people would *suppress* the invocation of the battery as an explanation for having a car-starting problem. What is equally obvious is that people do not store all these hypothetical conditions explicitly with each conceivable explanation of car-starting problems but treat them as unattached, *implicit suppressors*, namely, conditions which exert their influence only upon becoming actively believed and, when they do, would uniformly suppress *every* E-def rule having "car not starting" as its sole antecedent.

But if the list of suppressors is not prepared in advance, how do people distinguish a genuine suppressor from one in disguise. In other words, by what criterion could people discriminate between the suppressor "the starter is burned" and the candidate suppressor "I hear no motor sound"? Either of these two inspires strong belief in "the car won't start" and "I'll be late for the meeting"; yet, the burned-out starter is licensed to suppress the conclusion "the battery is dead," while the motor's silence is licensed to evoke it. I submit that it is in the *causal directionality* of the suppressor-suppressed relationship which provides the identification criterion: the antecedents of causal rules do qualify as suppressors while those of evidential rules do not. It is hard to see how implicit suppression could be realized, had people not been blessed with clear distinction between *explanation-evoking* and *expectation-evoking* rules. So, why stifle this distinction in formal reasoning systems?

Formally, implicit suppression can be defined in terms of a metarule that qualifies the viability of every E-def rule $P \rightarrow_E Q$ in the system:

$$P \rightarrow_E Q \,|\, \text{UNLESS } \exists(Q')\colon (Q' \rightarrow_C P) \text{ and } [E(Q') \text{ or } C(Q')]\,.$$

The rule says that the default rule $P \rightarrow_E Q$ can be invoked only when no alternative explanation Q' of P is believed. One may, in fact, turn things around and take this vulnerability to implicit suppression as the defining criterion for evidential rules. Accordingly, a rule $P \rightarrow Q$ will be said to be evidential if there exists another rule $Q' \rightarrow P$, such that Q' & P is a reason to believe $\neg Q$, while Q' alone is not a reason to believe $\neg Q$. Otherwise, the rule will be called causal.

The main benefit of this suppression scheme is that we no longer need to prepare the name of each potential suppressor next to that of a would-be suppressed; the connection between the two will be formed "on the fly," once the suppressor becomes actively believed. The mere fact that a belief in a proposition P can be justified by some explanation Q' would automatically and precisely block all the rules we wished suppressed. More ambitiously, it should also lead to retracting all conclusions drawn from premature activation of such rules as is demonstrated in Section 3. This is one of the computational benefits offered by the organizational instrument called causation. It has so far been realized using the numerical representation of Bayesian inference, but, since human reasoning is mostly qualitative, it would be interesting to embody in nonnumeric systems as well.

Unfortunately the benefit of implicit suppression is hindered by some fundamental issue, and it is not clear how it might be realized in purely categorical systems which preclude any representation for the degree of support that a premise imparts to a conclusion. Treating *all* C-def rules as implicit suppressors would be inappropriate, as was demonstrated in the starting theme of this note. In Case 1 of Joe's story, we correctly felt uncomfortable letting his father's profession inhibit the E-def rule

$$\text{CAN-READ(JOE)} \rightarrow_E \text{OVER-7(JOE)}\ ,$$

while now we claim that certain facts (e.g., burned starter), by virtue of having such compelling predictive influence over other facts (e.g., car not starting), should be allowed to inhibit all E-def rules emanating from the realization of such predictions (e.g., dead battery). Apparently there is a sharp qualitative difference between *strong* C-def rules such as

$$\text{HAS}(z, \text{BURNED-STARTER}) \rightarrow_C \text{WON'T-START}(z)$$

and *weak* C-def rules such as

$$\text{ENGLISH-PROFESSOR}(\text{FATHER}(z)) \rightarrow_C \text{CAN-READ}(z)$$

or

$$\text{HAS}(z, \text{OLD-STARTER}) \rightarrow_C \text{WON'T-START}(z)\ .$$

Strong C-def rules, if invoked, should inhibit all E-def rules emanating from their consequences. On the other hand, weak C-def rules should allow these E-def rules to fire (via rule (c)).

This distinction is exactly the role played by the numerical parameters in Bayesian inference; they measure the accrued strength of causal support, and serve to distribute the impact of newly observed facts among those propositions which had predicted the observations. Normally, those propositions which generated strong prior expectations of the facts observed would receive the major share of support imparted by the observation [8]. It is primarily due to this strong versus weak distinction that Bayesian inference rarely leads to counter-intuitive conclusions, and this is also why it is advisable to consult Bayes' analysis as a standard for abstracting more refined logical systems. However, the purpose of this note is not to advocate the merits of numerical schemes but, rather, to emphasize the benefits we can draw from the distinction between causal and evidential default rules. It is quite feasible that with just a rough quantization of rule strength, the major computational benefits of causal reasoning could be tapped.

5. Conclusion

The distinction between *C*-believed and *E*-believed propositions allows us to properly discriminate between rules that should be invoked (e.g., Case 1 of Joe's story) and those that should not (e.g., Case 2 of Joe's story), without violating the original intention of the rule provider. While the full power of this distinction can, admittedly, be unleashed only in systems that are sensitive to the relative strength of the default rules, there is still a lot that causality can offer to systems lacking this sensitivity.

ACKNOWLEDGMENT

I thank H. Geffner, V. Lifschitz, D. McDermott, P. Morris, J. Minker, D. Perlis and C. Petrie for their comments on an earlier version of this paper, and two anonymous reviewers for suggesting significant improvements.

This work was supported in part by the National Science Foundation, Grant DCR 8644931.

REFERENCES

1. Charniak, E., Riesbeck, C.K. and McDermott, D.V., *Artificial Intelligence Programming* (Erlbaum, Hillsdale, NJ, 1980).
2. Doyle, J., A truth maintenance system, *Artificial Intelligence* **12** (1979) 231–272.
3. Hanks, S. and McDermott, D., Nonmonotonic logic and temporal projection, *Artificial Intelligence* **33** (1987) 379–412.
4. Morris, P., Curing anomalous extensions, in: *Proceedings AAAI-87*, Seattle, WA (1987) 437–442.
5. Pearl, J., Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29** (1986) 241–288.
6. Pearl, J., Distributed revision of composite beliefs, *Artificial Intelligence* **33** (2) (1987) 173–215.
7. Pearl, J., A probabilistic treatment of the Yale shooting problem, Tech. Rept. R-100, Cognitive Systems Laboratory, UCLA (1987).
8. Pearl, J., Canonical models for causal interactions, Tech. Rept. R-104, Cognitive Systems Laboratory, Computer Science Department, UCLA (1987); also in: J. Pearl, *Networks of Belief: Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufman, Los Altos, CA, 1988) Ch. 4.
9. Pearl, J. and Verma, T., The logic of representing dependencies by directed graphs, in: *Proceedings AAAI-87*, Seattle, WA (1987) 374–379.
10. Poole, D.L., Defaults and conjectures: Hypothetical reasoning for explanation and prediction, Research Rept. CS-87-54, University of Waterloo, Waterloo, Ont. (1987).
11. Reiter, R. and Cricuolo, G., Some representational issues in default reasoning, *Int. J. Comput. Math.* **9** (1983) 1–13.
12. Schachter, R.D. and Heckerman, D., A backward view for assessment, *AI Mag.* **8** (3) (1987) 55–61.
13. Shortliffe, E.H., *Computer-Based Medical Consultation: MYCIN* (Elsevier, New York, 1976).

Received June 1987; revised version received January 1988