Notre Dame Journal of Formal Logic
Volume 49, Number 3, 2008

Causality, Modality, and Explanation

Graham White

Abstract We start with Fodor's critique of cognitive science in "The mind doesn't work that way: The scope and limits of computational psychology": he argues that much mental activity cannot be handled by the current methods of cognitive science because it is nonmonotonic and, therefore, is global in nature, is not context-free, and is thus not capable of being formalized by a Turing-like mental architecture. We look at the use of nonmonotonic logic in the artificial intelligence community, particularly with the discussion of the so-called frame problem. The mainstream approach to the frame problem is, we argue, probably susceptible to Fodor's critique; however, there is an alternative approach, due to McCain and Turner, which is, when suitably reformulated, not susceptible. In the course of our argument, we give a proof theory for the McCain-Turner system and show that it satisfies cut elimination. We have two substantive conclusions: first, that Fodor's argument depends on assumptions about logical form which not all nonmonotonic theories satisfy and, second, that metatheory plays an important role in the context of evolutionary accounts of rationality.

1 Introduction

Fodor [10] has argued that mental processes fall into two classes: those which are modular (that is, effectively encapsulated from other mental processes) and those which are not. The modular ones take place by means of local, syntactic operations on mental representations ([10], pp. 18–19): if they function properly, they will be insensitive to what may or may not happen outside their module. In fact, more than this is true: such processes are *context invariant* ([10], pp. 34–35). All they care about is the syntax of the mental representations which they are working on. In particular, they are monotonic: the validity of the inferences they perform is unaffected by adding extra mental representations to their context.

The nonmodular ones, by contrast, are not monotonic. Fodor's examples of such nonmodular processes are, first, abduction, and, second, the belief change apparatus

Received January 17, 2007; accepted January 21, 2008; printed June 6, 2008
2000 Mathematics Subject Classification: Primary, 03B45; Secondary, 03B45, 03F05
Keywords: explanation, causality, modal logic, cut elimination
 10.1215/00294527-2008-015

associated with abductive inference (including judgments as to the centrality of particular beliefs in one's belief system). These processes are, plausibly, nonmonotonic and context-sensitive ([10], pp. 33ff.); thus, Fodor goes on to argue, they cannot be described as operations on the (context-invariant) syntax of mental representations.

Now one *could* quibble about all this in various ways. One could worry about whether Fodor had distinguished sufficiently well between abstract and concrete syntax [23]. One could wonder whether he had taken sufficiently into account the possibility of context dependence which was regular enough not to have the deleterious consequences that he wishes to avoid [31]. But, for all that, his argument is cogent and worrying.

Rather than raising these sorts of objections to Fodor's argument, it may be more illuminating to look at how the artificial intelligence community deals with this sort of problem. We will look at what is described in the AI community as the "frame problem"; although this problem, which is mainly concerned with reasoning about change in the world, is somewhat different from the belief change problems that Fodor discusses, the two areas are systematically related [5], so that discussions of one should, mutatis mutandis, apply to the other.

Our results are interesting. We investigate two approaches to the frame problem. One of them, the more mainstream one, is based on the idea of minimizing change and probably *is* susceptible to critique in Fodor's manner. However, the other one, due to McCain and Turner, is probably not thus susceptible: it shows illuminatingly how reasoning can be nonmonotonic without being catastrophically global.

In arguing for this position, we have work, both technical and philosophical, to do. The McCain-Turner system is usually presented model-theoretically: we develop a proof theory for it, in order to show directly what reasoning in that system is like. In particular, we prove a cut elimination result, which can be regarded as showing how the system can support nonmonotonic inference and still be tractable. Proofs of these results are quite technical and have been modularized into the appendix of this paper.

We also have philosophical work to do. McCain and Turner present their system as a system of what they term "causal reasoning": I argue that the system can, much more plausibly, be presented as a logic of *explanation*. This, of course, brings it much closer to Fodor's concern with large-scale cognitive architecture, in which explanation ought to play a central role. So, in order to make this claim more plausible, I describe an interesting logic of argument due to Parsons and Jennings: this logic turns out to be merely a notational variant of a fragment of the McCain-Turner system.

2 The Logical Background

2.1 Basics Regardless of the details of the common approaches to reasoning about action, they all have several things in common. They all assume that one should start with a logical description of actions and their effects; they all (or mostly all) assume that the logical apparatus should deliver a description of the sequence of states of affairs resulting from a given sequence of actions. We will, for the sake of simplicity, assume that we have linear discrete time, indexed by the natural numbers: propositions will have an argument place for the time, into which terms standing for temporal instants can be substituted. Devotees of more sophisticated approaches—for example, of Reiter's elegant situation calculus formulation [28]—are welcome to translate my oversimplified notation into their variant.

2.2 The frame problem A fundamental difficulty behind using logic to reason about action in this way is what is known as the frame problem. The problem is that, given the state of the world before an action, there are many, merely "logically possible," states of the world after it, but most of these states are physically unrealistic. This remains so even if we give what seem to be perfectly adequate logical formalizations of actions.

Here is a famous example of what is at issue.

Example 2.1 (The Yale Shooting Problem) Suppose that we describe a state of the world at a particular time t by propositions alive and loaded. The world (a particularly simple one) contains a victim, a gun (and, presumably, a shooter, but he or she can remain unformalized). alive_t will say that some victim is still alive at t, whereas loaded_t will say that a gun is loaded at that time. Consider actions shoot and wait: wait has no effect, whereas shoot, if the gun is loaded, kills the victim. So we can specify the shoot action as follows (for the sake of simplicity, we assume that every action takes one unit of time).

$$\mathsf{shoot}_t \wedge \mathsf{loaded}_t \vdash \neg\mathsf{alive}_{t+1} . \tag{1}$$

Suppose now we have the following scenario: the gun is loaded at time 0, then there is a wait action, and the gun is then shot. So the following sequence of truth values is compatible with our specification.

t	0 wait	1 shoot	2
alive	$\top$	$\top$	$\bot$
loaded	$\top$	$\top$	$\bot$

(2)

Unfortunately, so is this.

t	0 wait	1 shoot	2
alive	$\top$	$\top$	$\top$
loaded	$\top$	$\bot$	$\bot$

(3)

However one might wish to diagnose the situation here, it is clear that something has gone wrong: logic, supplemented by unproblematic-seeming axiomatizations of the effects of actions, generates more solutions to these problems than it should. We need extra constraints, and there have been at least two attempts to say what sort of constraints we should have: a mainstream approach, based on the idea of minimization, and an alternative approach, rather less clearly formulated, but technically more viable and more interesting. As I shall argue, the alternative approach is, in fact, based on the idea of explanation.

2.3 Minimization-based approaches We first outline, and criticize, the minimization-based approaches. This will not only show some of the issues at stake, but it will also give an example of an approach which is, arguably, susceptible to Fodor's critique.

The basic idea behind these approaches is simple: that we should pick those histories which minimize the amount of change. One can, presumably, motivate this criterion as follows: the axiomatizations of actions give the changes which *must* take place if the actions are to be performed, and one wants exactly these changes and no others. Hence one should have a minimal amount of change. This idea is very close to Lewis's treatment of counterfactuals [14; 15; 16] and thus has a respectable philosophical pedigree. It can, furthermore, be supported by a mathematical argument: a wide class of nonmonotonic logics can be specified by giving a preference relation on models of some monotonic theory and then selecting only those models which are minimal according to the preference relation [6]. So this approach seems to have the advantage of a strong philosophical motivation, together with mathematical generality.

However, problems arise when one tries to make this approach more specific. Example 2.1 shows the first of these problems: there are two histories in that example, and the wrong one has one change of truth value, whereas the correct one has two changes. Even if we minimize by looking at containment relations between sets of changes, rather than naïvely counting them, the problem persists. So minimization often gives incorrect results. One can perhaps fix it, but many of the fixes (for example, preferring later changes to earlier ones) seem to be merely ad hoc.

The second problem is this: what does one mean by "amount of change"? This is admittedly a vague description, but it is not clear how to make it any less vague. The usual approach is this: one fixes a language, and then, for each history, one fixes the set of truth-value changes (that is, the set of ordered pairs consisting of a time and a primitive of the language which changes truth-value at that time). Containment between these sets gives a preference relation between histories: those histories are to be preferred which have minimal sets of truth-value changes. One can give this procedure a much more sophisticated and formal definition: it is generally known as *circumscription* [17] and is widely used in the AI community.

The problem with this approach is that it is strongly language-dependent: it is possible to have equivalent languages, with different primitives, which give different preference relations according to this definition. One can, by a suitable choice of primitives for the language, make the preference relation come out any way that one wants [33]. In more technical language, the entailment relations which circumscription yields are not preserved by uniform substitution. Failure of uniform substitution is (*pace* Makinson [19]) a cause of concern, for two reasons. Practically, it makes the use of these theories rather problematic: there may well be some magical set of primitives which make the predictions come out right, but one is given no guidance as to what it might be. And, theoretically, this remains a worry: closure under uniform substitution is motivated by a concern that mathematics should be more than mere syntax, that is, that the entities which mathematics and logic describe should be independent of our choice of the language—the "coordinate frame," as it were—which we use to describe them.

Now Lewis works with possible worlds, which for him are first-class individuals, and his closeness relation between possible worlds—which corresponds to our notion of minimizing change—is simply given: he does not say how to work it out, and, in particular, how to work it out on the basis of physically realistic measurements. Consequently, he does not explicitly face the problems that we have described above.

However, we should note that his account may well be susceptible to similar problems: it may work in the abstract, but if we require that the nearness relations between worlds should be capable of evaluation in terms of continuously differentiable functions of measurements made in those worlds, we may end up with a theory incompatible with physics [34]. Lewis's approach turns out to be, in Fodor's terms, *global*: we have to know about all of the possible worlds that there are, and we have to know these worlds directly as individuals in order to make causal inferences.

The difficulty with Lewis's account comes from a quite general result (specifically, from what is known as Noether's theorem) which has to do with the invariance of the laws of physics under coordinate transformation. So conceptually these considerations are not infinitely far removed from the problems that circumscription has with invariance under uniform substitution—and the problems with this approach seem to be quite deeply rooted.

It is, of course, not impossible that one could modify circumscription so as to allay these worries or, indeed, that Lewis's account of counterfactuals could be made specific enough to work with physical laws. However, these are difficult problems, and they may well be better illuminated by an alternative approach.

2.4 Explanation-based approaches Consider the bad history given in (3). We might attempt a critique of it as follows: the problem is not that there are *too many* changes, but that we cannot *explain* one of the facts—namely, the falsity of loaded at time 1. All of the other facts in that history can be explained, either because of the effects of actions or because they are facts unchanged from the previous time, but not this one. So a good constraint to put on our logic might be to require that every fact should be explained.

This approach might have a good deal to recommend it. A good notion of explanation would, one thinks, be language-, or coordinate-frame-, independent: the translation of an explanation would surely still be an explanation. Furthermore, it has a pleasing directness: human beings like us perform a lot of common sense causal explanation, and reflection on this process might give us useful data for formalizing such explanations. We might also cite the philosophical tradition: since the time of Aristotle, at least, causality and explanation have been very closely linked ([1], p. 32; [9]).

In order to carry this program forward, we will have to have a certain amount of technical apparatus. We will have to have a notion of explanation: the things to be explained will be facts, and we should be able to express, of any particular fact, whether it is explained or not. These notions should be closed under uniform substitution; that is, the translation of an explanation should be an explanation. Such an approach to reasoning about action is, indeed, possible; it was formulated by McCain and Turner, and we will study it in the remainder of this paper.

3 McCain and Turner's System

Their treatment [22] of the frame problem goes as follows. Suppose that we have a language $\mathfrak{L}$, which will have the connectives of the classical propositional calculus. McCain and Turner then consider a binary sentential operator; let us write it $\cdot \rhd \cdot$. Applications of this operator—of the form $\varphi \rhd \psi$—are called "causal rules," and they can, in McCain and Turner's treatment, be regarded as purely metatheoretic assertions: *they* read them as 'φ causes ψ', but we will argue for a reading of 'φ

explains ψ'. A collection of causal rules will be called a *causal theory*. Given a causal theory Θ in our language $\mathfrak{L}$, we define an operator $(\cdot)^{\Theta}$, from models of $\mathfrak{L}$ to sets of sentences, as follows:

$$M^{\Theta} \quad = \quad \overline{\{\psi | \varphi \rhd \psi \in D, M \Vdash \varphi\}}, \tag{4}$$

where $\bar{\cdot}$ is closure under entailment. We now say that a model M of $\mathfrak{L}$ is *causally explained* if and only if it is the only model of M^{Θ}. And we say that a proposition P is a *consequence* of Θ if and only if it is true in all of the Θ-causally explained models of $\mathfrak{L}$.

3.1 McCain and Turner's causal rules We will give a simplified version of McCain and Turner's causal theory. In order to make our exposition more direct, we will be presenting a version of our sequent calculus without the quantifiers (we could describe a version with quantifiers, but it would add extra technical complexity, and this extra complexity would not be germane to the main argument of this paper).

Theories of this sort will talk of two sorts of entities, *fluents* and *actions*: a fluent is something which can be true or false at a particular time, whereas an action is something which can be performed (or not) at a particular time. We represent each of these by a series of temporally indexed propositional atoms:

$$f_t \quad (t \in \mathbb{N}) \qquad \text{for each fluent,} \tag{5}$$

$$a_t \quad (t \in \mathbb{N}) \qquad \text{for each action.} \tag{6}$$

Intuitively, f_t is true if and only if f holds at time t, and a_t is true if and only if the action a occurs at time t. The atoms f_t will be called the *fluent atoms*: a *fluent literal* will be a fluent atom or the negation of a fluent atom.

We now describe our causal theory. We assume that the effects of the actions can be formalized by giving, for each action a, a finite set of *precondition-postcondition pairs*. Such a pair consists of a precondition and a postcondition, each of them a conjunction of fluent literals: the intuitive meaning of it is that, if the precondition is true in a situation, then, after the execution of the relevant action, the postcondition will be true in the successor situation. For example, the precondition of the shoot action is the loaded fluent, and the postcondition of the shoot action is $\neg\mathsf{alive} \wedge \neg\mathsf{loaded}$.

So, for each action a, and for each precondition-postcondition pair $\langle f, g\rangle$ belonging to that action, we add a series of causal rules:

$$(f_t \wedge a_t) \rhd g_{t+1} \tag{7}$$

where f is the precondition of a, and g is the postcondition. We also have to say which actions occur at which times: so, if action a occurs at time t_0, we add the causal rule

$$a_{t_0} \rhd a_{t_0}. \tag{8}$$

We also need to deal with fluents that remain unchanged (see [30]): so, for every fluent literal f_t, we add the causal rule

$$f_t \wedge f_{t+1} \rhd f_{t+1}. \tag{9}$$

Finally, we assume that a set of fluents is given which describes the initial situation: for each such fluent f, we add the causal rule

$$f_0 \rhd f_0. \tag{10}$$

Given these rules, this procedure appears to work; that is, it yields correct solutions of the frame problem (the reader is invited to check that it works with Example 2.1). As a bonus, it is fairly efficient computationally.

3.2 The road ahead As it stands though, McCain and Turner's definitions are not entirely perspicuous, either mathematically or conceptually. The definition of causally explained models is mysterious; we would, thus, like a more illuminating treatment of the mathematics. Furthermore, it is difficult to explain the conceptual role of $\triangleright$. For example, one of McCain and Turner's rules is $f_0 \triangleright f_0$ (10). But reading $\triangleright$ as 'causes' here is simply implausible: propositions—or the states of affairs which they denote—are not usually thought of as causing themselves.

So we will do two things in this paper. First, we will give an alternative definition of (4): it can easily be reformulated as a relation between models, and relations on a set of models give modal operators. So we can instead define a suitable modal operator $\Box$; given this operator, we can reformulate the definition of M^{Θ} (4) as

$$M^{\Theta} \quad = \quad \{\varphi | M \Vdash \Box\varphi\};$$

similarly, the relation

$$\varphi \vdash \Box\psi$$

is equivalent to

$$\varphi \triangleright \psi$$

(equivalent in the sense that the two relations give the same set of "causally closed" models).

But, as well as merely technical reformulations, we also want to say what these constructions *mean*. We argue then that a better reading of $\varphi \triangleright \psi$ is that 'φ explains ψ'—and, correspondingly, $\Box\psi$—can be regarded as a disjunction of all of the possible explanations of ψ. In fact, this explanatory reading has been anticipated: Lifschitz ([18], p. 451) paraphrases $\varphi \triangleright \psi$ as "[ψ] has a cause if [φ] is true, or ... [φ] provides a 'causal explanation' for [ψ]." And, given this explanatory reading, reflexive rules seem far less contentious: (4) simply says that the proposition f_0 "explains itself." Self-explanation is a good deal less problematic than self-causation, since explanations, after all, have to come to an end (or, of course, loop) at some point—see, for example, Wittgenstein [36], §217. In a similar way, McCain and Turner's rule for persistence (9) can be easily motivated in terms of explanation: a good explanation for a fluent being true at $t+1$ is that it was true at t and that its truth value is unchanged between t and $t+1$. But it is hard to read this rule *causally* without, again, invoking self-causation.

4 The Modal Logic

4.1 Motivation Our system is defined proof-theoretically, but it was arrived at by reformulating McCain and Turner's original definition. Our aim was to find a sequent calculus, with a good proof theory, which corresponded very closely to their model-theoretic construction. We proceed by progressively reformulating their work.

Remark 4.1 (Notation) We should explain a few notational conventions. φ and ψ will always stand for, respectively, the antecedents and consequents of causal rules. General propositions will be written using lowercase Roman letters, $p, q, r, \ldots$. Sets of propositions will be written with uppercase Greek letters, Γ, Δ; Θ will be used for causal theories, that is, sets of causal rules. Languages will be written $\mathfrak{L}$, with subscripts and superscripts. Models of our various theories will be written $M, M', \ldots$. And in the appendix when we do ordinal analysis we will use lowercase Greek letters, but not φ or ψ.

4.1.1 Reformulation 1: Model theory First we give a definition.

Definition 4.2 A causal theory Θ defines a relation $\mathbf{R}_\Theta$ on models of the language $\mathfrak{L}$ by

$$M\ \mathbf{R}_\Theta\ M' \Leftrightarrow \text{for every } \varphi \rhd \psi \in \Theta, \text{ if } M \Vdash \varphi, \text{ then } M' \Vdash \psi. \tag{11}$$

We then have the following lemma.

Lemma 4.3 *For any $p \in \mathfrak{L}$,*

$$M^\Theta \vdash p \ \Leftrightarrow\ M' \Vdash p \textit{ for all } M' \textit{ with } M\ \mathbf{R}_\Theta\ M'. \tag{12}$$

Proof Note first that, for any model M',

$$M' \Vdash M^\Theta \ \Leftrightarrow\ \text{for all } \varphi \rhd \psi \in \Theta.\, M \Vdash \varphi \Rightarrow M' \Vdash \psi;$$

so we have

$$\begin{aligned} M^\Theta \vdash p &\Leftrightarrow M' \Vdash p \text{ for all } M' \Vdash M^\Theta \\ &\Leftrightarrow \text{for all } M' \text{ (for all } \varphi \rhd \psi \in \Theta\ M \Vdash \varphi \Rightarrow M' \Vdash \psi) \\ &\qquad\qquad \Rightarrow M' \Vdash p \\ &\Leftrightarrow \text{for all } M'.M\ \mathbf{R}_\Theta\ M' \Rightarrow M' \Vdash p. \end{aligned}$$

□

So we have this corollary.

Corollary 4.4 *For any model M, M is causally explained according to Θ if and only if*

$$\{M\} = \{M' | M\ \mathbf{R}_\Theta\ M'\}.$$

4.1.2 Reformulation 2: Modal models We can now reformulate McCain and Turner's theory in modal terms.

Definition 4.5 Given, as above, a language $\mathfrak{L}$ and a causal theory Θ, define a Kripke model K_Θ as follows:

the language is $\mathfrak{L}_\Box$, generated by $\mathfrak{L}$ together with the modal operator $\Box$;

the frame consists of the set $\mathfrak{M}$ of models of the nonmodal language $\mathfrak{L}$, together with the accessibility relation $\mathbf{R}_\Theta$;

the forcing relation is given by the usual $\Vdash$ relation between elements of $\mathfrak{M}$ and propositions of $\mathfrak{L}$, extended to modal formulas in the usual way. If we wish to be pedantic (but we rarely will), we could call the new forcing relation $\Vdash_\Box$.

Notice that we have the following propositions.

Proposition 4.6 *For each causal law $P \rhd Q$, the modal sentence*

$$P \to \Box Q \tag{13}$$

is valid at each world of K.

Proof This follows directly from the definition of $\mathbf{R}_\Theta$. □

Proposition 4.7 *For $M \in \mathfrak{M}$, M is causally explained by Θ if and only if, as a world of the Kripke model K_Θ,*

$$M \Vdash (\Box p \to p) \wedge (p \to \Box p), \tag{14}$$

for any p.

Proof This is a standard modal reformulation of the condition

$$\{M\} = \{M' | M\ \mathbf{R}_\Theta\ M'\};$$

see, for example, van Benthem [3]. □

This result is pretty rather than useful: it refers to a particular Kripke model, and models are hard to present in any effective sense.

We can make progress though by asking how *other* Kripke models are related to K_Θ.

Definition 4.8 Let K be any other Kripke model of $\mathfrak{L}_\Box$ such that, for any $\varphi \rhd \psi \in \Theta$ and any world w of K,

$$w \Vdash \varphi \to \Box\psi.$$

Then define a map

$$\eta : \text{worlds of } K \quad \to \quad \text{worlds of } K_\Theta$$

by sending a world w of K to

$$\{p \mid p \in \mathfrak{L}, w \Vdash p\}. \tag{15}$$

Note that the sets of propositions defined by (15) are, by the properties of Kripke models, models of $\mathfrak{L}$; consequently, these sets of propositions are in fact worlds of K_Θ.

Proposition 4.9 *For w, w' worlds of K, if $w R_K w'$, then $\eta(w) R_\Theta \eta(w')$.*

Proof Suppose that $\varphi \rhd \psi \in \Theta$ and that $\eta(w) \Vdash \varphi$. Then, by definition of η, $w \Vdash \varphi$. However, by the assumption on K, $w \Vdash \varphi \to \Box\psi$; since $w R_K w'$, we have $w' \Vdash \psi$, and so, by the definition of η, $\eta(w') \Vdash \psi$. This is so for any $\varphi \rhd \psi \in \Theta$, and thus we have $\eta(w) R_\Theta \eta(w')$. □

Proposition 4.10 *For any world w of K and any nonmodal proposition p, if $\eta(w) \Vdash \Box p$, then $w \Vdash \Box p$.*

Proof Immediate from the above. □

This gives us an intuition as to what characterizes the model K_Θ: among all of the models which globally satisfy (13), it is the one in which □ is strongest (or, alternatively, the one in which the relation R is the largest). This shows us how to find a sequent calculus which corresponds to the model K_Θ, and which, consequently, represents the McCain-Turner procedure.

There is a standard way of getting such a sequent calculus (see, for example, Hallnäs and Schroeder-Heister [11; 12]). We first give ourselves right rules for $\Box$ corresponding to the constraints which $\Box$ must satisfy; then we find left rules which invert the right rules. The constraints which $\Box$ must satisfy are quite simple: $\varphi \to \Box\psi$ must be a theorem, for every $\varphi \rhd \psi \in \Theta$. Rather more generally, if we have causal rules $\{\varphi_i \rhd \psi_i\}_{i\in I}$ and if we have $\{\psi_i\}_{i\in I} \vdash p$, then we must have $\{\varphi_i\}_{i\in I} \vdash \Box p$; this gives us a right rule for $\Box$ which ensures that $\Box$ is a **K** modality. The left rule will, correspondingly, invert this right rule, and we get the rules for $\Box$ in Table 1. And, as the remainder of this paper will show, our system does indeed give a sequent calculus for the McCain-Turner procedure. It allows us to recover the Kripke model K_Θ, which turns out to be the canonical model of our modal logic.

4.2 The system Our system will be given by a sequent calculus as in White [35]; it is given by the rules in Table 1, whose formulation depends on an underlying set of causal rules. As explained above, we will for the sake of simplicity present the propositional version of this system. We could present a first-order system, but at the price of extra technical complexity.

We should note that this sequent calculus introduces a new language $\mathfrak{L}_\Box$—an extension of $\mathfrak{L}$ by the modal operator $\Box$—and a new consequence relation $\vdash_\Box$. Our cut elimination theorem will show that $\mathfrak{L}_\Box$ is a definitional extension of $\mathfrak{L}$ and that $\vdash$ is the restriction of $\vdash_\Box$ to $\mathfrak{L}$. But, for the moment, we will be careful to respect the differences between the two entailment relations.

Remark 4.11 The following considerations may make the calculus and the connection between $\mathfrak{L}_\Box$ and $\mathfrak{L}$ more perspicuous. If we were to assume that $\mathfrak{L}$ has arbitrary disjunctions, our cut elimination result would show that we could write

$$\Box p = \bigvee_{\psi_1,\ldots,\psi_k \vdash_\Box p} \varphi_i \wedge \cdots \wedge \varphi_k .$$

In this case, the left and right rules for $\Box$ would simply be the left and right rules for a disjunction $\bigvee_{i\in I} P_i$, namely,

$$\frac{\{\Gamma, p_i \vdash_\Box \Delta\}_{i\in I}}{\Gamma, \bigvee_{i\in I} p_i \vdash_\Box \Delta} \bigvee \mathrm{L} \qquad \text{and} \qquad \frac{\Gamma \vdash_\Box p_i, \Delta \quad (i \in I)}{\Gamma \vdash_\Box \bigvee_{i\in I} p_i} \bigvee \mathrm{R} . \tag{16}$$

This explains the form of the side conditions $\psi_1, \ldots, \psi_k \vdash_\Box p$ in the left and right rules for $\Box X$: they take the place of $i \in I$ in the left and right rules for $\bigvee_{i\in I}$. Just as in those latter left and right rules, the conditions $\psi_1, \ldots, \psi_k \vdash_\Box p$ occur positively in both left and right rules: they are side conditions and not premises.

Remark 4.12 We should remark that the rules for $\Box$ are, *in general*, infinitary (and even when finite, the set of premises of $\Box$L can well be undecidable). The qualification "in general" is important here: cut elimination will show that, for typical applications, this rule very often has a finite, decidable (and, indeed, very tractable) set of premises. Indeed, many applications—for example, that described in Appendix B—will only use the right rule for $\Box$, which is much more tractable. Furthermore, even when the rules are actually infinitary, the system is still very useful for metatheoretical purposes; after all, the proof-theoretic use of infinitary systems has a very long and respectable history (see, for example, Fairtlough and Wainer [8], p. 164).

Table 1 Sequent Calculus Rules

$$\frac{}{p \vdash_{\Box} p}\ \text{Ax}$$

$$\frac{}{\bot \vdash_{\Box}}\ \bot\text{L} \qquad \frac{}{\vdash_{\Box} \top}\ \top\text{R}$$

$$\frac{\Gamma \vdash_{\Box} \Delta}{\Gamma, p \vdash_{\Box} \Delta}\ \text{LW} \qquad \frac{\Gamma \vdash_{\Box} \Delta}{\Gamma \vdash_{\Box} p, \Delta}\ \text{RW}$$

$$\frac{\Gamma, p, p \vdash_{\Box} \Delta}{\Gamma, p \vdash_{\Box} \Delta}\ \text{LC} \qquad \frac{\Gamma \vdash_{\Box} p, p, \Delta}{\Gamma \vdash_{\Box} p, \Delta}\ \text{RC}$$

$$\frac{\Gamma \vdash_{\Box} p, \Delta}{\Gamma, \neg p \vdash_{\Box} \Delta}\ \neg\text{L} \qquad \frac{\Gamma, p \vdash_{\Box} \Delta}{\Gamma \vdash_{\Box} \neg p, \Delta}\ \neg\text{R}$$

$$\frac{\Gamma, p, q \vdash_{\Box} \Delta}{\Gamma, p \wedge q \vdash_{\Box} \Delta}\ \wedge\text{L} \qquad \frac{\Gamma \vdash_{\Box} p, \Delta \quad \Gamma \vdash_{\Box} q, \Delta}{\Gamma \vdash_{\Box} p \wedge q, \Delta}\ \wedge\text{R}$$

$$\frac{\Gamma, p \vdash_{\Box} \Delta \quad \Gamma, q \vdash_{\Box} \Delta}{\Gamma, p \vee p \vdash_{\Box} \Delta}\ \vee\text{L} \qquad \frac{\Gamma \vdash_{\Box} p, q, \Delta}{\Gamma \vdash_{\Box} p \vee q, \Delta}\ \vee\text{R}$$

$$\frac{\Gamma \vdash_{\Box} p, \Delta \quad \Gamma, q \vdash_{\Box} \Delta}{\Gamma, p \to q \vdash_{\Box} \Delta}\ \to\text{L} \qquad \frac{\Gamma, p \vdash_{\Box} q, \Delta}{\Gamma \vdash_{\Box} p \to q, \Delta}\ \to\text{R}$$

$$\frac{\Gamma \vdash_{\Box} \varphi_1 \wedge \cdots \wedge \varphi_k, \Delta \quad \psi_1, \ldots, \psi_k \vdash_{\Box} p}{\Gamma \vdash_{\Box} \Box p, \Delta}\ \Box\text{R}$$

$$\frac{\left\{\Gamma, \{\varphi_i\}_{i \in I_j} \vdash_{\Box} \Delta, \quad \{\psi_i\}_{i \in I_j} \vdash_{\Box} p\right\}_{j \in J}}{\Gamma, \Box p \vdash_{\Box} \Delta}\ \Box\text{L}$$

$$\frac{\Gamma \vdash_{\Box} p^m, \Delta \quad \Gamma', p^n \vdash_{\Box} \Delta'}{\Gamma, \Gamma' \vdash_{\Box} \Delta, \Delta'}\ \text{multicut}$$

Conditions on the rules

$\Box$R	where, for all i, $\varphi_i \rhd \psi_i$ is a causal rule.
$\Box$L	for each appropriate i, we have a causal rule $\varphi_i \rhd \psi_i$, and where the $\{\varphi_i \rhd \psi_i\}_{i \in I_j}$ are the only such finite sets of φs and ψs, such that $\{\psi_i\}_{i \in I_j} \vdash_{\Box} p$, that there are ($J$ need not be finite).
multicut	p^n stands for n occurrences of p; $m, n > 0$.

Example 4.13 Here is an example of how the left rule might plausibly be used. Scriven ([29], p. 61) describes the following argument pattern. We know that event p occurred. We know that the causes of p are a, b, or c, but we also know that neither a nor b occurred. So we conclude c.

In our system, this would go as follows. We have p, and we are supposing the world to be explanatorily closed, so we can conclude $\Box p$ (here we are doing causal explanation, so our explanations really are causes). We can also conclude that

$$\Box p \cong a \vee b \vee c;$$

we also have $\neg b$ and $\neg c$, so we conclude a.

5 Consequences of Cut Elimination

The proof of cut elimination is technical (it relies on ordinal induction) and is given in the Appendix. It has important consequences, however, and we give some of them here.

Corollary 5.1 *The modal theory is a conservative extension of the nonmodal theory; that is, if* Γ *and* Δ *are nonmodal, we have*

$$\Gamma \vdash \Delta \quad \textit{iff} \quad \Gamma \vdash_{\Box} \Delta.$$

Proof Any proof of $\Gamma \vdash \Delta$ is a proof of $\Gamma \vdash_{\Box} \Delta$; however, any cut-free proof of $\Gamma \vdash_{\Box} \Delta$ cannot involve the modal rules and is thus a proof of $\Gamma \vdash \Delta$. □

So from now on we can safely ignore the distinction between $\vdash$ and $\vdash_{\Box}$.

5.1 Semantics We can now prove a soundness and completeness theorem for our logic. We do this as follows: we first identify our model K_Θ of Definition 4.5 as the canonical model of this modal logic. That is, semantic validity in this single model will entail syntactic validity. We can then prove soundness by showing that any syntactically valid sequent is also semantically valid when interpreted in the canonical model and prove completeness by showing the converse. Recall that K_Θ has, as worlds, models of the nonmodal language $\mathfrak{L}$ and that the accessibility relation is given by the relation $\mathbf{R}_\Theta$ of Definition 4.2. The corresponding modeling relation between worlds of K_Θ and sentences of $\mathfrak{L}_\Box$ will, as in Definition 4.5, be written $\Vdash_\Box$.

For convenience of notation, we identify models of $\mathfrak{L}$ with the sets of nonmodal sentences true in them, and these are just the maximal $\vdash$-consistent sets of nonmodal sentences: let $\mathfrak{M}$ be the set of all such models. Clearly, for A nonmodal and $M \in \mathfrak{M}$, we have

$$M \Vdash a \quad \text{iff} \quad a \in M \quad \text{iff} \quad M \vdash a.$$

For this section, letters $a, a', \ldots$ will range over nonmodal sentences, whereas $p, p', \ldots$ will range over the whole of $\mathfrak{L}_\Box$.

The proof of the following is immediate.

Lemma 5.2 *For any* M *and any nonmodal* a, $M \Vdash_\Box a$ *if and only if* $M \Vdash a$.

We now define a semantic entailment relation, $\vDash_\Box$: this will be like the usual notion of modal semantic entailment, but with respect to the single model K_Θ. As we have explained, K_Θ will turn out to be the canonical model of our logic, so the restriction to K_Θ is harmless.

We first define semantic values for sentences in $\mathfrak{L}_\Box$.

Definition 5.3 For sentences p in $\mathfrak{L}_\Box$, define their *semantic values* $[\![p]\!] \subseteq \mathfrak{M}$ as follows:

$$\begin{array}{rcll} [\![a]\!] & = & \{M | M \Vdash a\} & \text{for } a \text{ a nonmodal atom} \\ [\![\neg p]\!] & = & \mathfrak{M} - [\![p]\!] & \\ [\![p \wedge q]\!] & = & [\![p]\!] \cap [\![q]\!] & \\ [\![p \vee q]\!] & = & [\![p]\!] \cup [\![q]\!] & \\ [\![\Box p]\!] & = & \{M | \forall M'. M \, \mathbf{R}_\Theta \, M' \Rightarrow M' \in [\![q]\!]\}. & \end{array}$$

The following lemma is immediate.

Lemma 5.4 *For a nonmodal,* $[\![a]\!] = \{M \in \mathfrak{M} | M \Vdash a\}$.

Definition 5.5 For $\Gamma, \Delta \subseteq \mathfrak{L}_\Box$, we say that

$$\Gamma \vDash_\Box \Delta \quad \text{iff} \quad \bigcap_{p \in \Gamma} [\![p]\!] \subseteq \bigcup_{q \in \Delta} [\![q]\!].$$

We also need to generalize our operation $(\cdot)^\Theta$ of (4) so that it can be applied to general subsets, rather than simply models, of $\mathfrak{L}$. We recall that, for $M \in \mathfrak{M}$, $M^\Theta \overset{\text{def}}{=} \overline{\{\psi | M \Vdash \varphi \text{ for some } \varphi \rhd \psi\}}$.

Definition 5.6 Let $S \subseteq \mathfrak{L}$: define

$$S^\Theta \overset{\text{def}}{=} \overline{\{a \in \mathfrak{L} \mid S \vdash \varphi_1 \vee \cdots \vee \varphi_k \text{ for rules } \varphi_i \rhd \psi_i, \text{ with } \psi_i \vdash a \text{ for all } i\}}.$$

Note that, if S is a model, the new definition coincides with the old one. We need the disjunction of φs in order to prove the following lemma, which will be crucial.

Lemma 5.7 *If* $\{S_i\}_{i \in I}$ *is a set of subsets of* $\mathfrak{L}$, *with each* S_i *closed under entailment, we have*

$$\bigcap_{i \in I} (S_i^\Theta) \quad = \quad \left(\bigcap_{i \in I} S_i \right)^\Theta .$$

Proof The right to left containment is trivial. For the left to right containment, suppose that $a \in S_i^\Theta$ for all i. Thus, for all i, we have

$$S_i \vdash \varphi_{i_1}, \ldots, \varphi_{i_{k_i}} \quad \text{with} \quad \psi_{i_j} \vdash a \; \text{ for all } \; j.$$

Then, since the S_i are closed under entailment, we have

$$\bigcap_{i \in I} S_i \vdash \bigcup_{i \in I} \{\varphi_{i_1}, \ldots, \varphi_{i_{k_i}}\}.$$

But, by the compactness of $\mathfrak{L}$, we can find a finite set of $\{\varphi_i \rhd \psi_i\}$ with

$$\bigcap_{i \in I} S_i \vdash \varphi_1, \ldots, \varphi_n \quad \text{with} \quad \psi_i \vdash a \; \text{ for all } \; i,$$

and this establishes the result. □

We now prove soundness and completeness.

Lemma 5.8 *The* K_Θ *semantics is sound and complete for inferences of the form* $\Gamma \vdash_\Box \Box p$, *for* $\Gamma \subseteq \mathfrak{L}$ *and* $p \in \mathfrak{L}_\Box$; *that is,*

$$\Gamma \vdash_\Box p \quad \textit{iff} \quad \Gamma \vDash_\Box p.$$

Proof We prove this by induction on the complexity of p: for p nonmodal, it is, by Corollary 5.1, simply soundness and completeness for the predicate calculus. The inductive steps for the nonmodal connectives are quite standard: there remains the case when $p = \Box q$. It will suffice to show that

$$[\![\Box q]\!] \quad = \quad \bigcup_i [\![\varphi_{i,1} \wedge \cdots \wedge \varphi_{i,k_i}]\!]$$

where, in the usual way, $\varphi_{i,j} \rhd \psi_{i,j}$ for all relevant i, j, where $\psi_{i,1}, \ldots, \psi_{i,k_i} \vdash_{\Box} q$ for all i, and where these are all of such finite sets of causal laws that there are; the right to left inference gives soundness, and the left to right inference gives completeness. Note that the semantic values of $\Box q$, for any q, are a union of the semantic values of nonmodals, and so

$$[\![\Box q]\!] = \bigcup_{[\![a]\!] \subseteq [\![\Box q]\!]} [\![a]\!] \, .$$

Now, if a is nonmodal,

$$\begin{array}{rcl} [\![a]\!] & \subseteq & [\![\Box q]\!] \\ & \text{iff} & \forall M . M \Vdash a . \forall M' . M' \Vdash M^{\Theta} \;\; \Rightarrow \;\; M' \Vdash_{\Box} q , \end{array}$$

expanding the definitions of $[\![\Box q]\!]$ and of $\mathbf{R}_{\Theta}$

$$\begin{array}{rl} \text{iff} & \forall M' . (\exists M \Vdash q . M' \Vdash M^{\Theta}) \;\; \Rightarrow \;\; M' \Vdash_{\Box} q \\ \text{iff} & \forall M' . M' \Vdash \left(\overline{\{A\}}\right)^{\Theta} \;\; \Rightarrow \;\; M' \Vdash_{\Box} q \end{array}$$

(where $\overline{\{a\}}$ is the deductive closure of a) by Lemma 5.7

$$\text{iff} \quad \left(\overline{\{a\}}\right)^{\Theta} \Vdash_{\Box} q$$

by the inductive hypothesis. Now expansion of the definitions shows that

$$\bigcup_{\left(\overline{\{a\}}\right)^{\Theta} \Vdash_{\Box} Q} [\![a]\!] \quad = \quad \bigcup_i [\![\varphi_{i,1} \wedge \cdots \wedge \varphi_{i,k_i}]\!]$$

(with the usual restrictions on the $\varphi_{i,j}$), which was to be proved. □

Recall now that the forcing relation, $\Vdash_{\Box}$, on the Kripke model is defined inductively using the structure of K_{Θ}. We can now show that, in fact, it is the same as provability.

Corollary 5.9 *For a model $M \in \mathfrak{M}$ and $p \in \mathfrak{L}_{\Box}$,*

$$M \Vdash_{\Box} p \quad \textit{iff} \quad M \vdash_{\Box} p .$$

Proof By soundness and completeness,

$$M \vdash_{\Box} p \quad \text{iff} \quad [\![M]\!] \subseteq [\![p]\!];$$

but $[\![M]\!] = \{M\}$, so

$$\begin{array}{rcl} M \vdash_{\Box} p & \text{iff} & M \in [\![p]\!] \\ & \text{iff} & M \Vdash_{\Box} p \end{array}$$

by definition of $\Vdash_{\Box}$ □

Theorem 5.10 *The K_Θ semantics is sound for general sequents.*

Proof We prove this by induction over the length of a proof. The nonmodal rules are straightforward, and we use Lemma 5.8 for the modal rules. □

Lemma 5.11 *For any $p \in \mathfrak{L}_\Box$, and for any $M \in \mathfrak{M}$, either $M \vdash_\Box p$ or $M \vdash_\Box \neg p$.*

Proof We prove this by induction. Since $\vdash_\Box$ coincides with $\vdash$ on $\mathfrak{L}$, and since M is a model of $\mathfrak{L}$, the result is clearly true if p is nonmodal. The crucial step is to prove the result for propositions of the form $\Box q$. Here we argue as follows:

$$\begin{array}{ll} \text{Suppose} & M \not\vdash_\Box \neg \Box q; \\ \text{then} & M, \Box q \not\vdash_\Box \bot, \end{array}$$

and, by the left rule for $\Box$, this holds if and only if, for some set of causal rules I,

$$M, \{\varphi_i\}_{i \in I} \not\vdash_\Box \bot$$

which, since M is a model of $\mathfrak{L}$, in turn holds if and only if, for any $i \in I$,

$$M \vdash_\Box \varphi_i,$$

and so, by the right rule for $\Box$,

$$M \vdash_\Box \Box q\,.$$

By cut elimination, no element of $\mathfrak{L}_\Box$ can be both entailed and not entailed by M. □

Corollary 5.12 *The maximal $\vdash_\Box$-consistent subsets of $\mathfrak{L}_\Box$ are in one-to-one correspondence with the maximal $\vdash$-consistent subsets of $\mathfrak{L}$ (i.e., the elements of $\mathfrak{M}$).*

Proof Intersection with $\mathfrak{L}$ gives a morphism in one direction. Lemma 5.11 shows that every maximal $\vdash$-consistent subset of $\mathfrak{L}$ can be uniquely extended to a maximal $\vdash_\Box$-consistent subset of $\mathfrak{L}_\Box$. □

Theorem 5.13 *The K_Θ semantics is complete.*

Proof It suffices to show that, if $\Gamma \subseteq \mathfrak{L}_\Box$, and if $\Gamma \not\vdash_\Box \bot$, then $[\![\Gamma]\!]$ is not empty, that is, that there is some world M of K_Θ with $M \in [\![\Gamma]\!]$. But, if we have $\Gamma \not\vdash_\Box \bot$, we can extend Γ to a maximal $\vdash_\Box$-consistent subset of $\mathfrak{L}_\Box$: this will correspond to an element M of $\mathfrak{M}$. By Corollary 5.9, this is the world we want. □

5.1.1 Interpolation and finiteness We first prove the following interpolation theorem.

Proposition 5.14 (Interpolation) *If we have $\Gamma \vdash_\Box \Box p, \Delta$, for some p and $\Gamma, \Delta \subseteq \mathfrak{L}_\Box$, then there is a set $\Delta_0 \subseteq \mathfrak{L}$ such that*

$$\Gamma \vdash_\Box \Delta_0, \Delta$$

and, for each $a \in \Delta_0$, there is some $b \in \mathfrak{L}$ such that

$$\begin{array}{ll} a \vdash_\Box \Box b & \textit{and} \\ b \vdash_\Box p. & \end{array}$$

Proof We prove this by a straightforward induction over a cut-free proof. Because we might use the contraction rule on $\Box p$, the inductive hypothesis will be the same as that of the theorem but with the slightly weaker condition that $\Gamma \vDash_{\Box} (\Box p)^n, \Delta$. The elements a of Δ_0 will turn out to be finite conjunctions of rule bodies, whereas the bs will be the corresponding conjunctions of rule heads. $\Box$

One of the key applications of our cut elimination theorem is in proving finiteness results such as the following. It is applicable to the modal implications that we will use in the next section for interpreting Parsons and Jennings' logic of argument, so it is quite significant.

Corollary 5.15 *If*

$$\Gamma \vdash \Box p$$

is provable in the sequent calculus, with Γ a set of nonmodals, then there is a finite set $\Gamma' \subseteq \Gamma$, and nonmodal propositions A and B, such that

$$\Gamma' \vdash a,$$
$$a \vDash_{\Box} \Box b, \quad \textit{and}$$
$$b \vDash_{\Box} p.$$

Proof We apply Theorem 5.14, which gets us a possibly infinite set Δ with $\Gamma \vDash_{\Box} \Delta$. But, in this case, Γ and Δ are both classical, so we can apply Corollary 5.1 to show that $\Gamma \vdash \Delta$. We can now apply compactness to find a *finite* $\Delta_0 \subseteq \Delta$, and, taking disjunctions, we get a single a and b. $\Box$

6 Explanation and Argument

As we have argued, McCain and Turner's theory seems to be a very natural formulation of explanation *in general*: although its original application may have been to a causal context, there is nothing about it which forces these explanations to be *causal*. Once we broaden our horizons to general explanation, we can bring this modal system into contact with other work: we give one example of this, the theory of Parsons and Jennings [24] (see also Parsons et al. [25]). We could give other applications: for example, where we have explanations, questions must also be in the neighborhood, and thus we can also express a good deal of the formalism of Belnap and Steel [2] (see Harrah [13]) in terms of ours.

6.1 The Parsons and Jennings system Parsons and Jennings ([24]; see also [25]) have described a consequence relation, $\vdash_{\text{ACR}}$, which is intended to capture the practice of argumentation. The items which this system manipulates are ordered pairs (p, Γ), where p is a proposition and A a set of propositions: we will call such a pair an *argument*. Intuitively, p is the conclusion of an argument, and Γ is the set of its grounds. We will also call p the *head* of the argument, and Γ its *body*.

The system is given in Table 2: here Θ is a set of *basic* arguments, and it plays exactly the same role in their system as a causal theory does in McCain and Turner's. Parsons and Jennings write their system in natural deduction style, with introduction and elimination rules, using sequents of the form $\Theta \vdash_{\text{ACR}} (p, A)$. Such a sequent says that A is an argument for p, given basic arguments Θ.

(Note that we have interchanged the labels on the rules $\neg$I and RAA. Parsons and Jennings' original labeling is clearly a typo of some sort.)

Table 2 The Parsons and Jennings System

$$\frac{(p,\Gamma)\in\Theta}{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)}\text{Axiom}\qquad\qquad \frac{}{\Theta\vdash_{\mathrm{ACR}}(\top,\varnothing)}\top\mathrm{I}$$

$$\frac{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)\quad\Theta\vdash_{\mathrm{ACR}}(q,\Gamma')}{\Theta\vdash_{\mathrm{ACR}}(p\wedge q,\Gamma\cup\Gamma')}\wedge\mathrm{I}\qquad\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(p\wedge q,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)}\wedge\mathrm{E1}\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(p\wedge q,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(q,\Gamma)}\wedge\mathrm{E2}$$

$$\frac{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p\vee q,\Gamma)}\vee\mathrm{I1}\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(q,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p\vee q,\Gamma)}\vee\mathrm{I2}\qquad\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(p\vee q,\Gamma)\quad \Theta,(p,\Gamma)\vdash_{\mathrm{ACR}}(r,\Gamma')\quad \Theta,(q,\Gamma)\vdash_{\mathrm{ACR}}(r,\Gamma'')}{\Theta\vdash_{\mathrm{ACR}}(r,\Gamma'\cup\Gamma'')}\vee\mathrm{E}$$

$$\frac{\Theta,(p,\varnothing)\vdash_{\mathrm{ACR}}(\bot,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(\neg p,\Gamma)}\neg\mathrm{I}\qquad\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)\;\Theta\vdash_{\mathrm{ACR}}(\neg p,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(\bot,\Gamma)}\neg\mathrm{E}$$

$$\frac{\Theta,(p,\varnothing)\vdash_{\mathrm{ACR}}(q,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p\to q,\Gamma)}\to\mathrm{I}\qquad\qquad \frac{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)\;\Theta\vdash_{\mathrm{ACR}}(p\to q,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(q,\Gamma\cup\Delta)}\to\mathrm{E}$$

$$\frac{\Theta\vdash_{\mathrm{ACR}}(\bot,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)}\mathrm{EFQ}\qquad\qquad \frac{\Theta,(\neg p,\varnothing)\vdash_{\mathrm{ACR}}(\bot,\Gamma)}{\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)}\mathrm{RAA}$$

6.2 Comparison with our system We can now translate Parsons and Jennings' system into ours.

Definition 6.1 Let $\Theta\vdash_{\mathrm{ACR}}(p,\Gamma)$ be a sequent in Parsons and Jennings' system. Its *modal translation* is the sequent

$$\Gamma\vdash\Box_\Theta P,$$

where $\Box_\Theta$ is the modal operator defined by causal laws

$$\left\{\left(\bigwedge_{q\in\Gamma} q\right)\rhd p \;\middle|\; (p,\Gamma)\in\Theta\right\}.$$

Since the Parsons and Jennings system is written in natural deduction style, some of the rules (for example, $\to$ I) manipulate the set of basic arguments; consequently, the modal operator in the modal translation will vary. We will, then, need the following lemma.

Lemma 6.2 *If $\Box$ is the modality associated to a set Θ of explanations and if $\Box'$ is that associated to $\Theta \cup \{(\psi, \varphi)\}$, then define an interpretation*

$$\alpha : \mathfrak{L}_{\Box'} \to \mathfrak{L}_{\Box}$$

by

$$\begin{aligned} \alpha(p) &= p && \textit{for } p \textit{ atomic} \\ \alpha(p \wedge p') &= \alpha(p) \wedge \alpha(p') \\ \alpha(p \vee p') &= \alpha(p) \vee \alpha(p') \\ \alpha(\neg p) &= \neg\alpha(p) \\ \alpha(\Box' p) &= (\varphi \wedge \Box(\psi \to \alpha(p))) \vee \Box\alpha(p) \end{aligned}$$

for any p. Then, for any $\Gamma, \Delta \subseteq \mathfrak{L}_{\Box'}$,

$$\Gamma \vdash_{\Box'} \Delta \quad \textit{iff} \quad \alpha(\Gamma) \vdash_{\Box} \alpha(\Delta).$$

Proof We check that Γ and $\alpha(\Gamma)$ have the same semantic values (regarded as subsets of $\mathfrak{M}$), and similarly for Δ; we then apply the soundness and completeness theorems. $\Box$

We have the following proposition.

Proposition 6.3 *The modal interpretation is* sound*; that is, each of Parsons and Jennings' axioms is translated into a tautology.*

Proof

Ax is

$$\frac{}{\Theta \vdash_{\mathrm{ACR}} (p, \Gamma)} \; (p, \Gamma) \in \Theta$$

and this follows from our definition of the modal translation.

∧−I, ¬−E, →−E $\wedge$−I, for example, is

$$\frac{\Theta \vdash_{\mathrm{ACR}} (p, \Gamma) \quad \Theta \vdash_{\mathrm{ACR}} (q, \Gamma')}{\Theta \vdash_{\mathrm{ACR}} (p \wedge q, \Gamma \cup \Gamma')}$$

and this follows from the **K** tautology $\Box p \wedge \Box q \vdash \Box p \wedge q$. $\neg$−E and $\to$−E are similar.

∧−E1, ∧−E2, ∨−I1, ∨−I2, EFQ $\wedge$−E1, for example, is

$$\frac{\Theta \vdash_{\mathrm{ACR}} (p \wedge q, \Gamma)}{\Theta \vdash_{\mathrm{ACR}} (p, \Gamma)}$$

and this follows from the **K** tautology $\vdash \Box(a \wedge b) \to \Box a$. $\wedge$−E2, $\vee$−I1, $\vee$I2, and EFQ are similar.

⊤−I This is just

$$\vdash \Box\top,$$

a **K** tautology.

∨−E This is

$$\frac{\Theta \vdash_{\mathrm{ACR}} (p \vee q, \Gamma) \quad \Theta, (p, \Gamma) \vdash_{\mathrm{ACR}} (r, \Gamma') \quad \Theta, (q, \Gamma) \vdash_{\mathrm{ACR}} (r, \Gamma'')}{\Theta \vdash_{\mathrm{ACR}} (r, \Gamma' \cup \Gamma'')}$$

and this corresponds, in our system, to

$$a \to \Box_{\Theta}(p \vee q),\ b \to \Box_{\Theta \cup \{(p,\Gamma)\}} r,\ c \to \Box_{\Theta \cup \{(q,\Gamma)\}} r$$
$$\vdash b \wedge c \to \Box_{\Theta} r.$$

We can use the lemma to express $\Box_{\Theta \cup \{(p,\Gamma)\}}$ and $\Box_{\Theta \cup \{(q,\Gamma)\}}$ in terms of $\Box_{\Theta}$; some routine but tedious computation then reduces this case to

$$\Box_{\Theta}(p \vee q), \Box_{\Theta}(p \to r), \Box_{\Theta}(q \to r) \vdash \Box_{\Theta} r, \tag{17}$$

which is a **K** tautology.

RAA, ¬−I ¬−I is

$$\frac{\Theta, (p, \varnothing) \vdash_{\mathrm{ACR}} (\bot, A)}{\Theta \vdash_{\mathrm{ACR}} (\neg p, A)}$$

which corresponds to

$$A \to \Box_{\Theta \cup \{(p,\varnothing)\}} \bot \vdash A \to \Box_{\Theta} \neg p;$$

using the lemma on $\Box_{\Theta \cup \{(p,\varnothing)\}}$, and some computation, reduces this to

$$\Box \bot \vee \Box (p \to \bot) \vdash \Box \neg p,$$

which is a **K** tautology. ¬−I is similar.

→ −I This is

$$\frac{\Theta, (p, \varnothing) \vdash_{\mathrm{ACR}} (q, A)}{\Theta \vdash_{\mathrm{ACR}} (p \to q, A)}$$

which corresponds to

$$A \to \Box_{\Theta \cup \{(p,\varnothing)\}} q \vdash A \to \Box_{\Theta}(p \to q).$$

The usual moves reduce this to

$$\Box q \vee \Box(p \to q) \vdash \Box(p \to q),$$

again a **K** tautology. □

Completeness does not hold. This is for trivial reasons: all rules (except ∨E) of the Parsons and Jennings system leave the body of the argument intact. A trivial induction on the length of proofs will yield the following.

Proposition 6.4 *In any proof of*

$$\Theta \vdash_{\mathrm{ACR}} (p, \Gamma),$$

A must be a union of the bodies of rules in Θ.

Since the modal sequent calculus certainly does not satisfy this condition, we cannot hope for completeness. What we need to do is to be able to compose proofs in the Parsons and Jennings system with natural deduction proofs for the grounds of an argument. We could, theoretically, write down another set of rules for doing this. However, we only need one extra rule, which is this.

Definition 6.5 Let *classical* ∨E be the following rule:

$$\frac{\Gamma \vdash q \vee r \quad \Theta \vdash_{\mathrm{ACR}} (r, \Gamma' \cup \{q\}) \quad \Theta \vdash_{\mathrm{ACR}} (r, \Gamma'' \cup \{r\})}{\Theta \vdash_{\mathrm{ACR}} (r, \Gamma \cup \Gamma' \cup \Gamma'')} \vee \mathrm{EC}$$

where $A \vdash q \vee r$ is an entailment in classical natural deduction.

We clearly have this proposition.

Proposition 6.6 *The modal translation is sound for classical or-elimination.*

And we can also prove completeness.

Theorem 6.7 *The modal translation is complete; that is, given a proof of*

$$\Gamma \vdash \Box_\Theta p, \tag{18}$$

there is a proof, in the Parsons and Jennings system together with classical or-elimination, of

$$\Theta \vdash_{\mathrm{ACR}} (p, \Gamma). \tag{19}$$

Sketch of proof We establish the following lemma.

Lemma 6.8 *If p is nonmodal, given a sequent calculus proof of*

$$\psi_1, \ldots, \psi_k \vdash p,$$

then there is a Parsons and Jennings proof of

$$\Theta \vdash (p, \{\varphi_1, \ldots, \varphi_k\}).$$

This lemma can be proved by first transforming the sequent calculus proof to a natural deduction proof and then observing that the Parsons and Jennings rules mirror the rules of classical natural deduction.

So now we can prove the theorem. We take a proof of $A \vdash \Box_\Theta p$, and from it derive a finite set I such that

$$\Gamma \vdash \bigvee_{i \in I} \varphi_{i_1} \wedge \cdots \wedge \varphi_{i_{k_i}}, \tag{20}$$

$$\Theta \vdash_{\mathrm{ACR}} (p, \{\varphi_{i_1}, \ldots, \varphi_{i_{k_i}}\}) \qquad \text{for any } i. \tag{21}$$

We then use classical or-elimination in order to glue together (20) and (21). □

Remark 6.9 As we see here, the natural deduction formulation is actually quite ambiguous as to what its premises are: in a proof of

$$\Theta \vdash_{\mathrm{ACR}} (p, \Gamma),$$

are the premises the basic arguments Θ, or the grounds Γ for the argument which is to be established? Now the system is set up as if the premises are the set Θ of basic arguments, and this gives a sense of composition of arguments which is valid; that is, from $\Theta \vdash_{\mathrm{ACR}} (p, \Gamma)$ and $\Theta', (p, \Gamma) \vdash_{\mathrm{ACR}} (q, \Gamma')$, we can establish $\Theta', \Theta \vdash_{\mathrm{ACR}} (q, \Gamma')$. But this is not enough: we *also* want to regard the grounds of arguments as premises.

The situation is clearly two-dimensional in something like Pratt's sense—he defines the *dimension* of a logic to be "the smallest number of variables and constants of the logic sufficient to determine the remaining variables and constants" [27]: the modal operator can be varied quite independently of the classical connectives, merely by altering the set of causal rules. Consequently, a formalism such as Masini's [20; 21] may well be more appropriate.

Remark 6.10 This translation between sequent calculus and the Parsons and Jennings natural deduction is, in addition, not very sensitive to the structure of proofs on either side: natural deduction proofs tend to transform the conclusion of the argument quite extensively before coming down to basic arguments. Sequent calculus proofs, by contrast, leave the conclusion unchanged until an application of $\Box R$, after which the proof is a matter of standard classical logic. In addition, the Parsons and Jennings system only represents a fragment of the full sequent calculus (namely, the entailments in which $\Box$ only occurs on the right). A natural deduction formulation of the entire sequent calculus would be interesting but would have to extend the Parsons and Jennings system quite considerably.

7 Conclusion: Two Approaches to Nonmonotonic Reasoning

There are two approaches to nonmonotonic reasoning. There is the generally accepted one, which may well be susceptible to Fodor's critique. It can be summarized in Brewka's words:

> To formalize human commonsense reasoning something different [from classical logic] is needed. Commonsense reasoning is frequently not monotonic. In many situations we draw conclusions which are given up in the light of further information. ([6], p. 2)

According to this view, nonmonotonic logic is applicable *globally*: it applies to *all* of our common sense reasoning. There is, so to speak, a small fragment of our reasoning which happens to be monotonic, but common sense is nonmonotonic by default. There would, then, be a large amount of reasoning which was both nonmonotonic and global. This, if true, would be fairly catastrophic. There has been quite a lot of recent success in implementing nonmonotonic logic, but it is still true that, precisely because of this globality, performance scales quite badly [7]. So, if we are supposed to do it on a very large scale, it cannot be expected to perform well.

Contrast this with the McCain-Turner system. There are two aspects that we need to examine: first, reasoning within the system itself, and, second, the relation between the set of causal rules and the modal operator. The system itself is monotonic. It is a modal logic of a fairly well-known sort, and it has a sequent calculus which admits cut elimination; we can, then, search for proofs efficiently. The dependence of the modal operator on the set of rules $\varphi \rhd \psi$ is, however, nonmonotonic: if we add new rules, thus changing the modal operator, then we may invalidate modal inferences that we previously made. However, only the *left* rule for $\Box$ is thus sensitive; the right rule is not. Consequently, although our system *is* nonmonotonic, it is so in a quite limited way: it is not susceptible to the sort of "anything might entail anything else" worries that Fodor has.

There is a final remark to be made. Many of the tractability results for our system do not follow directly from its definition, but are established on the basis of metatheory, and quite technical metatheory at that. Now this would, perhaps, be a fatal objection if we were supposed to adopt this system by means of introspection. However, if we are supposed to acquire styles of reasoning on the basis of natural selection—in the manner of the cognitive science described in [10], Ch. 5—then none of this matters; the metatheoretical results apply to the system and give it the evolutionary advantages that it arguably has, regardless of whether or not the early primates who, maybe, adopted these styles of reasoning were, or were not, familiar with technical results in proof theory.

In the Appendix, we have material that does not fit naturally into the argumentative structure of the main paper: the cut elimination, which is too technical, and some material on Turner's logic of universal causation, which relates what we have done to previous work.

Appendix A Cut Elimination

We now prove our cut elimination result. This needs a certain amount of machinery, because our system is, in general, infinitary: proof trees may, therefore, be infinitely branching, and—since we cannot apply Zorn's lemma—there may be branches of infinite length above a given node. Because of this, we cannot assign a finite depth to each node, and, consequently, we cannot prove cut elimination in the usual manner, that is, by an induction on both the depth and the complexity of the cut formula. We can use an induction, but it must be an induction over ordinals: there is a machinery of ordinal analysis, which is used for the proof-theoretic analysis of infinitary systems (for example, arithmetic with the ω-rule). We will generally follow the treatment in Pohlers [26], with the modifications necessary because our system is two-sided, not one-sided, and because our sole infinitary connective is $\Box$ rather than the infinite conjunctions and disjunctions which Pohlers investigates. Our case, in fact, is rather simple, because, if we regard $\Box$ as a possibly infinite disjunction of finite conjuncts of nonmodal formulas, the complexity of the disjuncts is bounded a priori by ω.

In this section, since we want to treat entailments uniformly, we will write rule applications as follows:

$$\frac{\{\Gamma_i \vdash_{\Box} \Delta_i\}_{i \in I}}{\Gamma \vdash_{\Box} \Delta}$$

where I can have zero, one, or two members (for the standard rules) or an arbitrary number (for $\Box$L).

We first define the *rank* of a formula.

Definition A.1 If $F \in \mathfrak{L}_{\Box}$, let its rank $\mathrm{rk}(F)$ be the ordinal defined inductively as follows:

1. $\mathrm{rk}(F) = 0$ if F is atomic.
2. $\mathrm{rk}(\neg F) = \mathrm{rk}(\forall x.F) = \mathrm{rk}(\exists x.F) = \mathrm{rk}(F) + 1$.
3. $\mathrm{rk}(F_1 \wedge F_2) = \mathrm{rk}(F_1 \vee F_2) = \mathrm{rk}(F_1 \to F_1) = \max(\mathrm{rk}(F_1), \mathrm{rk}(F_2)) + 1$.
4. $\mathrm{rk}(\Box F) = 0$.

Note that $\mathrm{rk}(\Box F)$ is independent of F. This is because, as we have remarked, we can give an a priori bound on the disjuncts that $\Box F$ can be thought of as expanding into. We have the following proposition.

Proposition A.2 *For any* $F \in \mathfrak{L}_{\Box}$, $\mathrm{rk}(F) < \omega$.

Proof Immediate. □

We define the following classification of inference rules.

Definition A.3 The *finitary* rules will be all of the rules apart from $\Box$L. The *noncut* rules will be all of the rules except cut.

Next, we define entailment symbols annotated with both the rank of cut formulas and the depth of the proof tree. There are two classes of inference rules: the finitary rules, and $\Box$L. $\Box$L is the only infinitary rule; our proof trees, then, will consist of portions where finitary rules are applied, separated by applications of $\Box$L. We will treat the two kinds of rule application differently, and we will, therefore, use two ordinals for measuring the depth of the tree. ζ will measure the number of applications of $\Box$L above the root of the tree which have proofs involving cuts above them; it will, in general, be infinite. α will measure the number of finitary rule applications above the root of the tree, but will be reset to zero after each application of $\Box$L; it will be finite. ρ will measure the rank of cut formulas and will likewise be reset to zero after any application of $\Box$L; it will, therefore, be finite. The last two rules will make the annotations fit smoothly with our inductive arguments, which will first recurse over ζ, then over ρ, then over α.

Definition A.4 Define the entailment relations $\vdash_{\rho}^{\alpha,\zeta}$ as follows.

Case 1 If $\Gamma \vdash_{\Box} \Delta$ is an instance of Ax, of $\perp$ L, or of $\top$R, then

$$\Gamma \vdash_{0}^{0,0} \Delta .$$

Case 2 If we have an instance of a finite noncut inference rule

$$\frac{\{\Gamma_i \vdash_{\Box} \Delta_i\}_{i \in I}}{\Gamma \vdash_{\Box} \Delta}$$

and if, for all $i \in I$, $\Gamma_i \vdash_{\rho}^{\alpha_i,\zeta} \Delta_i$ where, for all i, $\alpha_i < \alpha$, then

$$\Gamma \vdash_{\rho}^{\alpha,\zeta} \Delta .$$

Case 3 If

$$\frac{\{\Gamma_i \vdash_{\Box} \Delta_i\}_{i \in I}}{\Gamma \vdash_{\Box} \Delta}$$

is an instance of $\Box$L, and if, for all i, $\Gamma_i \vdash_{0}^{\alpha_i,0} \Delta_i$ where $\zeta_i < \zeta$, then

$$\Gamma \vdash_{0}^{0,0} \Delta .$$

Case 4 If

$$\frac{\{\Gamma_i \vdash_{\Box} \Delta_i\}_{i \in I}}{\Gamma \vdash_{\Box} \Delta}$$

is an instance of $\Box$L, and if, for all i, $\Gamma_i \vdash_{\rho_i}^{\alpha_i,\zeta_i} \Delta_i$ where either some $\zeta_i > 0$ or some $\rho_i > 0$ and where $\zeta_i < \zeta$ for all i, then

$$\Gamma \vdash_{0}^{0,\zeta} \Delta .$$

Case 5 If

$$\Gamma \vdash_{\rho}^{\alpha_1,\zeta} A, \Delta \quad \text{and} \quad \Gamma', A \vdash_{\rho}^{\alpha_2,\zeta} \Delta'$$

with $\alpha_1, \alpha_2 < \alpha$ and $\mathrm{rk}(A) < \rho$, then

$$\Gamma, \Gamma' \vdash_{\rho}^{\alpha,\zeta} \Delta, \Delta' .$$

Case 6 If $\Gamma \vdash_{\rho}^{\alpha,\zeta} \Delta$, and if $\alpha \leq \beta, \rho \leq \sigma$ with β and σ finite, then

$$\Gamma \vdash_{\sigma}^{\beta,\zeta} \Delta .$$

Case 7 If $\Gamma \vdash^{\alpha,\zeta}_{\rho} \Delta$, and if $\zeta < \eta$, then, for any finite β and σ,

$$\Gamma \vdash^{\beta,\eta}_{\sigma} \Delta.$$

The following lemmas follow by a trivial induction.

Lemma A.5 *If $\Gamma \vdash_{\Box} \Delta$, then, for some finite α and ρ, and some ordinal ζ,*

$$\Gamma \vdash^{\alpha,\zeta}_{\rho} \Delta.$$

Lemma A.6 *A proof $\Gamma \vdash_{\Box} \Delta$ is cut free if and only if it has an annotation of the form*

$$\Gamma \vdash^{\alpha,0}_{0} \Delta.$$

Theorem A.7 *Given a McCain-Turner causal theory, the corresponding modal system satisfies cut elimination.*

We prove this by the following lemmas. They are very much the same as the corresponding lemmas in [26], pp. 60ff. Note that $\alpha\#\beta$ stands for the so-called natural sum of the ordinals α and β ([26], p. 43). The relevant facts that we shall use are that it is defined for any pair of ordinals α and β, and that it is strictly monotonic in *both* arguments.

Lemma A.8 *If*

$$\Gamma \vdash^{\alpha,\zeta}_{\rho} X, \Delta \quad \textit{and} \quad \Gamma', X \vdash^{\beta,\eta}_{\rho} \Delta'$$

and if $\mathrm{rk}(X) = \rho$*, then*

$$\Gamma, \Gamma' \vdash^{\alpha+\beta,\zeta\#\eta}_{\rho} \Delta, \Delta'.$$

Proof We prove this by induction, first on $\zeta\#\eta$, then on $\alpha+\beta$. There are four cases.

Case 1 One of the premises is an axiom, $\bot$L, or $\top$R, and X is not principal in it. In this case, $\Gamma, \Gamma' \vdash_{\Box} \Delta, \Delta'$ is an instance of the same rule, and, by the definition of $\Gamma, \Gamma' \vdash^{\alpha+\beta,\zeta\#\eta}_{\rho} \Delta, \Delta'$, we have the first case of the result.

Case 2 X is nonprincipal in one or the other of the premises (suppose, without loss of generality, the left one), and that premise is the conclusion of a finitary rule. Then the prooftree on the left looks like

$$\frac{\left\{\begin{array}{c}\Pi_i\\ \vdots\\ \Gamma_i \vdash^{\alpha_i,\zeta}_{\rho} X, \Delta_i\end{array}\right\}_{i\in I}}{\Gamma \vdash^{\alpha,\zeta}_{\rho} X, \Delta}\,.$$

Since, for all i, $\alpha_i < \alpha$ (and, consequently, $\alpha_i + \beta < \alpha + \beta$), we can assume the hypothesis inductively for cuts between $\Gamma_i \vdash^{\alpha_i,\zeta_1}_{\rho} X, \Delta_i$ and $\Gamma', X \vdash^{\beta,\zeta_2}_{\rho} \Delta'$, obtaining proofs Π_i, we complete the proof as follows:

$$\frac{\left\{\begin{array}{c}\tilde{\Pi}_i\\ \vdots\\ \Gamma_i, \Gamma' \vdash^{\alpha_i+\beta,\zeta\#\eta}_{\rho} \Delta_i, \Delta'\end{array}\right\}_{i\in I}}{\Gamma, \Gamma' \vdash^{\alpha+\beta,\zeta\#\eta}_{\rho} \Delta, \Delta'}\,.$$

Case 3 X is nonprincipal (without loss of generality, on the left), the bottom sequent on the left is the conclusion of $\Box$R, all of the premises on the left have $\zeta = \rho = 0$, and the sequent on the right has $\eta = \rho = 0$. So, the prooftree on the left looks like

$$\frac{\left\{\begin{array}{c}\Pi_i\\ \vdots\\ \Gamma_i \vdash_0^{\alpha_i,\,0} X, \Delta_i\end{array}\right\}_{i\in I}}{\Gamma \vdash_0^{\alpha,\,0} X, \Delta}\,.$$

We can apply the lemma inductively, obtaining proofs of

$$\Gamma_i, \Gamma' \vdash_0^{\alpha_i+\beta,\,0} \Delta_i, \Delta';$$

we now complete the proof as follows,

$$\frac{\left\{\begin{array}{c}\tilde{\Pi}_i\\ \vdots\\ \Gamma_i, \Gamma' \vdash_0^{\alpha_i+\beta,\,0} \Delta_i, \Delta'\end{array}\right\}_{i\in I}}{\Gamma, \Gamma' \vdash_0^{0,\,0} \Delta, \Delta'}\,.$$

Case 4 X is nonprincipal (without loss of generality, on the left), the bottom sequent on the left is the conclusion of $\Box$R, and we do not have the previous case. So, the prooftree on the left looks like

$$\frac{\left\{\begin{array}{c}\Pi_i\\ \vdots\\ \Gamma_i \vdash_{\rho_i}^{\alpha_i,\,\zeta_i} X, \Delta_i\end{array}\right\}_{i\in I}}{\Gamma \vdash_{\rho}^{\alpha,\,\zeta} X, \Delta}$$

where, for all i, $\zeta_i < \zeta$. We can apply the lemma inductively, obtaining proofs of $\Gamma_i, \Gamma' \vdash_{\rho_i'}^{\alpha_i+\beta,\,\zeta_i+\eta} \Delta_i, \Delta'$, where, for each i, $\zeta_i' \leq \zeta_i < \zeta$, and where $\rho_i' = \max(\rho_i, \eta)$: we now complete the proof as follows

$$\frac{\left\{\begin{array}{c}\tilde{\Pi}_i\\ \vdots\\ \Gamma_i, \Gamma' \vdash_{\rho_i'}^{\alpha_i',\,\zeta_i'} \Delta_i, \Delta'\end{array}\right\}_{i\in I}}{\Gamma, \Gamma' \vdash_0^{0,\,\zeta} \Delta, \Delta'}$$

and apply Case 6 of Definition A.4 to adjust the values of α and ρ.

Case 5 X is principal on both sides; there are various cases, depending on the principal connective. We treat two of these cases; the others are very similar.

($\to$) The proofs are

$$\frac{\begin{array}{c}\Pi_1\\ \vdots\\ \Gamma, Y \vdash_{\rho}^{\alpha_1,\,\zeta} Z, \Delta\end{array}}{\Gamma \vdash_{\rho}^{\alpha,\,\zeta} Y \to Z, \Delta} \to \mathrm{R} \qquad \text{and} \qquad \frac{\begin{array}{c}\Pi_2\\ \vdots\\ \Gamma' \vdash_{\rho}^{\beta_1,\,\eta} Y, \Delta'\end{array} \qquad \begin{array}{c}\Pi_3\\ \vdots\\ \Gamma', Z \vdash_{\rho}^{\beta_2,\,\eta} \Delta'\end{array}}{\Gamma', Y \to Z \vdash_{\rho}^{\beta,\,\eta} \Delta'} \to \mathrm{L}$$

with $\alpha_1 < \alpha$ and $\beta_1, \beta_2 < \beta$. We transform this into

$$\dfrac{\begin{array}{c}\tilde{\Pi}\\ \vdots\\ \Gamma, \Gamma' \vdash_{\rho}^{\alpha_1+\beta_1,\, \zeta\#\eta} Z, \Delta, \Delta'\end{array} \qquad \begin{array}{c}\Pi_3\\ \vdots\\ \Gamma', Z \vdash_{\rho}^{\beta_2,\, \eta} \Delta'\end{array}}{\Gamma, \Gamma' \vdash_{\rho}^{\alpha+\beta,\, \zeta\#\eta} \Delta, \Delta'}$$

where $\tilde{\Pi}$ is produced from Π_1 and Π_2 by the inductive hypothesis. We have $\mathrm{rk}(Y), \mathrm{rk}(Z) < \rho$, $\alpha_1\#\beta_1 < \alpha\#\beta$, and $\beta_1 < \alpha\#\beta$, so the final inference is justified.

($\Box$) The proofs are

$$\dfrac{\begin{array}{c}\Pi\\ \vdots\\ \Gamma \vdash_{\rho}^{\alpha_1,\, \zeta} \varphi_1 \wedge \cdots \wedge \varphi_k, \Delta\end{array}}{\Gamma \vdash_{\rho}^{\alpha,\, \zeta} \Box Y, \Delta}\ \Box\mathrm{R} \quad \text{and} \quad \dfrac{\left\{\begin{array}{c}\Pi_i'\\ \vdots\\ \Gamma', \varphi_{i_1}, \ldots, \varphi_{i_{k_i}} \vdash_{\rho}^{\beta_i,\, \eta_i} \Delta'\end{array}\right\}_{i\in I}}{\Gamma', \Box Y \vdash_{\rho}^{\beta,\, \eta} \Delta'}\ \Box\mathrm{L}$$

where $\alpha_1 < \alpha$, $\eta_i < \eta$ for all i.

We construct a new proof as follows. The tuple $\varphi_1, \ldots, \varphi_k$ must match one of the tuples indexed by I; say it matches i_0. So we now have proofs

$$\begin{array}{c}\Pi\\ \vdots\\ \Gamma \vdash_{\rho}^{\alpha_1,\, \zeta} \varphi_1 \wedge \cdots \wedge \varphi_k, \Delta\end{array} \quad \text{and} \quad \begin{array}{c}\Pi_{i_0}'\\ \vdots\\ \dfrac{\Gamma', \varphi_1, \ldots, \varphi_k \vdash_{\rho}^{\beta_{i_0},\, \eta_{i_0}} \Delta'}{\Gamma', \varphi_1 \wedge \varphi_2, \ldots, \varphi_k \vdash_{\rho}^{\beta_{i_0}+1,\, \eta_{i_0}} \Delta'}\ \wedge\mathrm{L}\,.\\ \vdots\ \wedge\mathrm{L}\\ \Gamma', \varphi_1 \wedge \cdots \wedge \varphi_k \vdash_{\rho}^{\beta_{i_0}+k-1,\, \eta_{i_0}} \Delta'\end{array}$$

We now have a proof of

$$\Gamma, \Gamma' \vdash_{\rho'}^{\alpha',\, \eta_{i_0}} \Delta, \Delta',$$

and $\eta_{i_0} < \eta$, so we can now apply Case 7 of Definition A.4 and obtain the result. □

Proof of Theorem A.7 Suppose that we have a proof involving cuts, that is, a proof of a sequent

$$\Gamma \vdash_{\rho+1}^{\alpha,\, \zeta} \Delta. \tag{22}$$

We prove that this proof can be replaced by a cut-free proof of the same sequent by an induction. The inductive hypothesis will be that, given such a proof, there is a cut-free proof

$$\Gamma \vdash_{0}^{\alpha',\, 0} \Delta$$

with some α'. We perform the induction first on ζ, then on ρ, then on α. There are four cases.

Case 1 If the last inference is an application of $\Box$L, then the premises are of the form $\Gamma_i \vdash_{\rho_i}^{\alpha_i',\, \zeta_i} \Delta_i$ with, for all i, $\zeta_i < \zeta$. We can assume the result inductively, obtaining cut-free proofs of the premises, and we can complete the proof as follows:

$$\frac{\{\Gamma_i \vdash_0^{\alpha''_i, 0} \Delta_i\}_{i \in I}}{\Gamma \vdash_0^{0,0} \Delta} \Box\text{L},$$

which is a cut-free proof of (22).

Case 2 If the last inference is a cut of rank less than ρ, then the premises will have the same value of ζ and ρ, but smaller values of α; inductively, we can assume that these premises have cut-free proofs. So we now have a proof as follows:

$$\frac{\Gamma' \vdash_0^{\alpha', 0} X, \Delta' \quad \Gamma'', X \vdash_0^{\alpha'', 0} \Delta''}{\Gamma', \Gamma'' \vdash_{\rho'}^{0,0} \Delta', \Delta''}$$

with $\rho' < \rho + 1$. So, by our inductive hypothesis, we can assume the result.

Case 3 If the last inference is a cut of rank ρ, then its premises will be of the form

$$\Gamma' \vdash_{\rho+1}^{\alpha', \zeta} \Delta' \quad \text{and} \quad \Gamma'' \vdash_{\rho+1}^{\alpha'', \zeta} \Delta''.$$

We can apply the result inductively to obtain cut-free proofs of the premises, which (after applying Definition A.4, Case 6) can be assumed to be of the form

$$\Gamma' \vdash_{\rho}^{\alpha', 0} \Delta' \quad \text{and} \quad \Gamma'' \vdash_{\rho}^{\alpha'', 0} \Delta'';$$

we then apply Lemma A.8 to obtain a proof of

$$\Gamma \vdash_{\rho}^{\alpha+\alpha', 0} \Delta.$$

We have possibly reduced the value of ζ, and certainly reduced the value of ρ, so we can assume the result by the inductive hypothesis.

Case 4 If the last inference is an application of a noncut finitary rule, then its premises must have the same value of ζ and ρ, but smaller values of α, than the conclusion. We can again apply the result inductively to the premises, and then apply the rule to the premises, giving a cut-free proof of (22). □

Appendix B Turner's *Logic of Universal Causation*

As we have indicated, Turner [32] also has a modal system which he uses as a metatheory for the McCain-Turner procedure. His treatment has the following features:

1. We are given an **S5** modal operator, written **C**, and a theory T in the language of that operator.
2. Recall that a Kripke model of **S5** is
   (a) a set $\mathcal{W}$ of worlds,
   (b) a truth-functional forcing relation between worlds and propositions of the nonmodal language, and
   (c) an equivalence relation on the set of worlds.

   We then say that an **S5** Kripke model of the theory T is *causally explained* if
   (a) it has a single world w;

(b) if any model of T has a set of worlds $\mathcal{W}$ which is a superset of $\{w\}$, with a forcing relation extending that for w, then $\{w\}$ is a single equivalence class in the larger model. (Equivalently, the inclusion of w in $\mathcal{W}$ is what is called a *p-morphism*; see van Benthem [3; 4]).

3. Then we say that a proposition is *causally explained* if it is forced at the unique world in all causally explained models.

Now this definition of a causally explained model involves not simply a model, but all models which contain it (in some appropriate sense of 'contain'): it is not obvious that this definition can be replaced with one which only talks, in the standard way, about theoremhood in a single model. In fact, as we shall see, it cannot.

Example B.1 Consider a language with a single atom p, and two theories in that language:

$$T_1 = \{p \to \mathbf{C}p\}; \tag{23}$$

$$T_2 = \{p \to \mathbf{C}p, \neg p \to \mathbf{C}\neg p\}. \tag{24}$$

Then T_1 has a single causally explained model: its world forces p. T_2 has two causally explained models, one with a single world forcing p, and one with a single world forcing $\neg p$. Consequently, the proposition p is causally explained by T_1, but not by T_2.

This has, as a consequence, the following.

Proposition B.2 *There is no set of propositions Γ such that, for any* **S5** *theory T,*

$$p \text{ is causally explained by } T \quad \Leftrightarrow \quad \Gamma, T \vdash_{\mathbf{S5}} p.$$

Proof If there were such a set of propositions Γ, then the relation

$$p \text{ is causally explained by } T$$

would be monotonic in T; as Example B.1 shows, it is not. □

By contrast, our logic is monotonic: it is given by a sequent calculus of the normal sort, *with* the weakening rule.

Example B.3 We should consider the analogue of Example B.1 in our system. The two theories would correspond to two different modal operators, $\Box_1$ and $\Box_2$, and we have

$$\Box_1 p \cong p \qquad \Box_1 \neg p \cong \bot$$
$$\Box_2 p \cong p \qquad \Box_2 \neg p \cong \neg p.$$

The two sets of propositions Γ_i are as follows:

$$\begin{aligned}\Gamma_1 &= \{\Box_1 p \leftrightarrow p, \Box_1 \neg p \leftrightarrow \neg p\}\\ &= \{p \leftrightarrow p, \neg\bot \leftrightarrow \neg p\}\\ &= \{\top, p\}.\\ \Gamma_2 &= \{\Box_2 p \leftrightarrow p, \Box_2 \neg p \leftrightarrow \neg p\}\\ &= \{p \leftrightarrow p, \neg p \leftrightarrow \neg p\}\\ &= \{\top\}.\end{aligned}$$

This is in accordance with the expected results: for T_1, only the world which forces p is causally explained, and so Γ_1 (essentially) contains only p. For T_2, both worlds are causally explained (both the one which forces p and the one which forces $\neg p$), and Γ_2 is essentially vacuous.

So, although our modal logic and Turner's yield the same sets of causally explained models (and hence the same sets of causally explained propositions), they have radically different properties. Ours is monotonic and has a well-behaved proof theory; Turner's is nonmonotonic and, as it stands, does not have a proof theory. Although Turner's system is formulated using **S5**, Proposition B.2 shows that we cannot, in any obvious way, use the proof theory of **S5** as a proof theory for his system.

References

[1] Barnes, J., *Aristotle*, Oxford University Press, London, 1982. 317

[2] Belnap, N. D., Jr., and T. B. Steel, Jr., *The Logic of Questions and Answers*, Yale University Press, New Haven, 1976. Zbl 0345.02015. MR 0536640. 328

[3] van Benthem, J., "Correspondence theory," pp. 167–247 in *Handbook of Philosophical Logic, Vol. 2*, edited by D. Gabbay and F. Guenther, vol. 165 of *Synthese Library*, Reidel, Dordrecht, 1984. Zbl 0875.03048. MR 844599. 321, 340

[4] van Benthem, J., *A Manual of Intensional Logic*, vol. 1 of *CSLI Lecture Notes*, Center for the Study of Language and Information, Stanford, 1985. Zbl 0678.03010. MR 867084. 340

[5] Bochman, A., *A Logical Theory of Nonmonotonic Inference and Belief Change*, Springer-Verlag, Berlin, 2001. Zbl 0977.03001. MR 1880387. 314

[6] Brewka, G., *Nonmonotonic Reasoning: Logical Foundations of Commonsense*, vol. 12 of *Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press, Cambridge, 1991. Zbl 0723.68099. MR 1115897. 316, 333

[7] Cadoli, M., F. M. Donini, and M. Schaerf, "Is intractability of non-monotonic reasoning a real drawback?" pp. 946–51 in *Proceedings of the Twelfth National Conference on Artificial Intelligence, Vol. 1, 2 (Seattle, 1994)*, AAAI Press, Menlo Park, 1994. citeseer.ist.psu.edu/article/cadoli96is.html. Zbl 0907.68182. MR 1301592. 333

[8] Fairtlough, M., and S. S. Wainer, "Hierarchies of provably recursive functions," pp. 149–207 in *Handbook of Proof Theory*, vol. 137 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1998. Zbl 0961.03053. MR 1640327. 322

[9] Falcon, A., "Aristotle on causality," The Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/archives/spr2006/entries/aristotle-causality, 2006. 317

[10] Fodor, J. A., *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, The MIT Press, Cambridge, 2000. 313, 314, 333

[11] Hallnäs, L., and P. Schroeder-Heister, "A proof-theoretic approach to logic programming. I. Clauses as rules," *Journal of Logic and Computation*, vol. 1 (1990), pp. 261–83. Zbl 0723.68028. MR 1152923. 322

[12] Hallnäs, L., and P. Schroeder-Heister, "A proof-theoretic approach to logic programming. II. Programs as definitions," *Journal of Logic and Computation*, vol. 1 (1991), pp. 635–60. Zbl 0769.68014. MR 1152930. 322

[13] Harrah, D., "Questions," pp. 1–3 in *Routledge Encyclopedia of Philosophy*, vol. 8, Routledge, London, 1998. 328

[14] Lewis, D., "Causation," pp. 159–213 in *Philosophical Papers*, vol. 2, Oxford University Press, London, 1986. Originally published in Journal of Philosophy, vol. 70 (1973), pp. 556–67. 316

[15] Lewis, D., "Counterfactual dependence and time's arrow," pp. 32–66 in *Philosophical Papers*, vol. 2, Oxford University Press, London, 1986. Originally published in Noûs, vol. 13 (1979), pp. 455–76. 316

[16] Lewis, D., "Counterfactuals and comparative possibility," pp. 3–31 in *Philosophical Papers*, vol. 2, Oxford University Press, London, 1986. Originally published in the Journal of Philosophical Logic, vol. 2 (1973), pp. 418–46. 316

[17] Lifschitz, V., "Circumscription," pp. 297–352 in *Handbook of Logic in Artificial Intelligence and Logic Programming. Nonmonotonic Reasoning and Uncertain Reasoning, Vol. 3*, Oxford Science Publications, Oxford University Press, New York, 1994. Zbl 0804.03017. MR 1281426. 316

[18] Lifschitz, V., "On the logic of causal explanation," *Artificial Intelligence*, vol. 96 (1997), pp. 451–65. Zbl 0901.03022. MR 1605309. 319

[19] Makinson, D., "Ways of doing logic: What was different about AGM 1985?" *Journal of Logic and Computation*, vol. 13 (2003), pp. 3–13. Zbl 1043.03008. MR 1973446. 316

[20] Masini, A., "2-sequent calculus: A proof theory of modalities," *Annals of Pure and Applied Logic*, vol. 58 (1992), pp. 229–46. Zbl 0819.03045. MR 1191942. 332

[21] Masini, A., "2-sequent calculus: Intuitionism and natural deduction," *Journal of Logic and Computation*, vol. 3 (1993), pp. 533–62. Zbl 0793.03013. MR 1253870. 332

[22] McCain, N., and H. Turner, "Causal theories of action and change," pp. 460–65 in *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference (Providence, 1997)*, AAAI Press, Menlo Park, 1997. MR 1484760. 317

[23] McCarthy, J., "Towards an abstract science of computation," http://www-formal.stanford.edu/jmc/towards.html, 1962. 314

[24] Parsons, S., and N. R. Jennings, "Negotiation through argumentation: A preliminary report," in *Proceedings of the International Conference on Multi Agent Systems (ICMAS'96)*, Kyoto, 1996. 328

[25] Parsons, S., C. Sierra, and N. Jennings, "Agents that reason and negotiate by arguing," *Journal of Logic and Computation*, vol. 8 (1998), pp. 261–92. citeseer.nj.nec.com/parsons98agents.html. Zbl 0904.68061. MR 1642062. 328

[26] Pohlers, W., *Proof Theory: An Introduction*, vol. 1407 of *Lecture Notes in Mathematics*, Springer, Berlin, 1989. Zbl 0695.03024. MR 1026933. 334, 336

[27] Pratt, V., "A roadmap of some two-dimensional logics," pp. 149–62 in *Logic and Information Flow*, edited by J. van Eijck and A. Visser, Foundations of Computing Series, The MIT Press, Cambridge, 1994. Zbl 0876.03014. MR 1295065. 332

[28] Reiter, R., *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*, The MIT Press, Cambridge, 2001. Zbl 1018.03022. 315

[29] Scriven, M., "Explanations, predictions, and laws," pp. 51–74 in *Theories of Explanation*, Oxford University Press, London, 1988. 323

[30] Shanahan, M., *Solving the Frame Problem. A Mathematical Investigation of the Common Sense Law of Inertia*, The MIT Press, Cambridge, 1997. MR 1445066. 318

[31] Strachey, C., "Fundamental concepts of programming languages," Lecture notes at the International Summer School in Programming Languages (Copenhagen, August 1967), 1967. 314

[32] Turner, H., "A logic of universal causation," *Artificial Intelligence*, vol. 113 (1999), pp. 87–123. Zbl 0940.03031. MR 1724108. 339

[33] White, G. G., "Intensionality and circumscription," pp. 372–79 in *Proceedings of the 9th International Conference on Non-Monotonic Reasoning (NMR2002)*, edited by S. Benferhat and E. Giunchiglia, Toulouse, 2002. 316

[34] White, G., "Lewis, causality, and possible worlds," *Dialectica*, vol. 54 (2000), pp. 133–37. MR 1770070. 317

[35] White, G., "A modal formulation of McCain and Turner's theory of causal reasoning," pp. 211–22 in *Logics in Artificial Intelligence (8th European Conference, JELIA 2002)*, edited by S. Flesca, S. Greco, N. Leone, and G. Ianni, vol. 2424 of *Lecture Notes in Computer Science*, Springer, Berlin, 2002. Zbl 1013.03026. MR 2052799. 322

[36] Wittgenstein, L., *Philosophical Investigations*, 3d edition, edited by G. E. M. Anscombe and R. Rhees, Blackwell Publishers, Oxford, 2001. Zbl 1028.03003. MR1971335. 319

Department of Computer Science
Queen Mary, University of London
London E2 9QD
UNITED KINGDOM
graham@dcs.qmul.ac.uk
http://www.dcs.qmul.ac.uk/~graham