

# Causality and Minimal Change Demystified

**Maurice Pagnucco**

Computational Reasoning Group  
Department of Computing  
Division of ICS  
Macquarie University  
NSW, 2109, Australia  
morri@ics.mq.edu.au

**Pavlos Peppas**

Dept. of Business Administration  
University of Patras  
Patras 26500, Greece  
ppeppas@otenet.gr

## Abstract

The Principle of *Minimal Change* is prevalent in various guises throughout the development of areas such as reasoning about action, belief change and nonmonotonic reasoning. Recent literature has witnessed the proposal of several theories of action that adopt an explicit representation of *causality*. It is claimed that an explicit notion of causality is able to deal with the frame problem in a manner not possible with traditional approaches based on minimal change.

However, such claims remain untested by all but representative examples. It is our purpose here to objectively test these claims in an abstract sense; to determine whether an explicit representation of causality is capable of providing something that the Principle of Minimal Change is unable to capture. Working towards this end, we provide a precise characterisation of the limit of applicability of minimal change.

## 1 Introduction

The problem of reasoning about action and change has been one of the major preoccupations for artificial intelligence researchers since the inception of the field. One of the early tenets applied when reasoning about such phenomena was that *as little as possible should change in the world when performing an action*; what we might call the *Principle of Minimal Change*.<sup>1</sup> This principle is manifest in many guises: preferential-style systems [Shoham, 1988], persistence approaches [Krautz, 1986], circumscription [McCarthy, 1980], etc. Over the years, aspects of this principle have been called into question leading to a variety of suggested fixes: fixed versus variable predicates in circumscription, occluded fluents [Sandewall, 1989], frame fluents [Lifschitz, 1990], to name but a few. Moreover, in the more recent literature explicit representations of causality have found favour [Lin, 1995; McCain and Turner, 1995; 1997; Thielscher, 1997]. However, what is not clear—beyond some simple representative

---

<sup>1</sup>Although, one might be tempted to say that the Principle of Minimal Change is more general in scope.

examples—is the purchase afforded by explicitly representing causality over the more traditional minimal change approaches. *It is this imbalance that this paper seeks to redress in a clear and objective manner.* In fact, the results we present here have further reaching consequences, giving a rather lucid characterisation of the extent of applicability of minimal change. By this we mean that, given a framework for reasoning about action and change, it will be clear whether such a framework can be modelled by minimal change once certain properties of the framework can be established.

We achieve our aims through a correspondence between two formal systems which we call *dynamic systems* and *preferential models* respectively. Intuitively, the dynamic system is an abstract modelling of the dynamic domain under consideration (the behaviour of which we wish to reason about). Essentially, this abstract model captures the domain at hand by a result function  $\mathcal{R}(w, \alpha)$  which returns the states (of the domain) that could possibly result from the application of an action with direct effects  $\alpha$  (i.e., postconditions) at the initial state  $w$ . A preferential model on the other hand is a formal structure that encodes the Principle of Minimal Change in an abstract and quite general manner. With the aid of preferential models we are able to provide a precise characterisation of the class of dynamic systems that are amenable to theories of action based on minimal change; we call such dynamic systems *minimisable*. Having a precise characterisation of minimisable dynamic systems we can then examine whether theories of action adopting an explicit representation of causality, which we shall call *causal theories of action* are capable of forms of reasoning that cannot be captured by the Principle of Minimal Change; more precisely, we can examine whether causal theories of action are applicable *outside* the scope of minimisable dynamic systems. According to the results reported herein, the logic of action proposed by Thielscher [1997] is indeed applicable to non-minimisable dynamic systems, whereas, perhaps surprisingly, McCain and Turner's causal theory of action [McCain and Turner, 1995] has a range of applicability that is subsumed by the class of minimisable dynamic systems.

In the following section we introduce both dynamic systems and preferential models. Moreover, we state clearly the notion of *minimal change* that we shall adopt here. In Section 3 we examine the formal properties that the result function of a dynamic system must obey in order for it to be *min-*

*imisable*. Section 4 presents an analysis of some of the theories of action found in the literature. We end with a discussion and conclusions, including pointers to future work.

## 2 Dynamic Systems and Preferential Systems

As mentioned above, the main results in this paper will be achieved by demonstrating a correspondence between two formal systems. The first, called a *dynamic system*, is meant to serve as a general abstraction of domains (such as the *blocks world*, or the domain described by the *Yale Shooting Problem*, etc.), for which theories of action are designed to reason about. Our main interest shall be in the properties of the dynamic system’s result function. In particular, we shall formulate necessary and sufficient conditions under which the system’s result function can be characterised in terms of an appropriately defined *minimisation policy*. Minimisation policies are in turn encoded by our second formal system called a *preferential model*. Dynamic systems and preferential models are formally defined below.

### 2.1 Dynamic Systems

Throughout this article we shall be working with a *finitary* propositional language  $\mathcal{L}$  the details of which shall be left open.<sup>2</sup> We shall often refer to  $\mathcal{L}$  as the *object language*. We shall call the propositional variables of  $\mathcal{L}$  *fluents*. The set of all fluents will be denoted by  $\mathcal{F}_{\mathcal{L}}$ . A *literal* is either a fluent or the negation of a fluent. We shall denote the set of all literals by  $\mathcal{Z}_{\mathcal{L}}$ . A *state* of  $\mathcal{L}$  (also referred to as an *object state*) is a maximally consistent set of literals. The set of all states of  $\mathcal{L}$  is denoted by  $\mathcal{M}_{\mathcal{L}}$ . For a set of sentences  $G$  of  $\mathcal{L}$ , by  $[G]$  we denote the set of all states of  $\mathcal{L}$  that satisfy  $G$ , i.e.  $[G] = \{r \in \mathcal{M}_{\mathcal{L}} : r \vdash G\}$ . Finally, for a sentence  $\varphi$  of  $\mathcal{L}$  we shall use  $[\varphi]$  as an abbreviation for  $\{[\varphi]\}$ .

**Definition 2.1** A dynamic system is a triple  $W = \langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$  where,

- $\mathcal{S}$  is a nonempty subset of  $\mathcal{M}_{\mathcal{L}}$  the elements of which we shall call *valid states*.
- $\mathcal{A}$  is a nonempty set of sentences of  $\mathcal{L}$ . The intended meaning of the sentences in  $\mathcal{A}$  is that they correspond to the *postconditions* (or *direct effects*) of actions. For simplicity, we shall identify actions with their *postconditions* and refer to the sentences in  $\mathcal{A}$  as *actions*.
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow 2^{\mathcal{S}}$ , is called the *result function*.

Intuitively, the result function  $\mathcal{R}(w, \alpha)$  returns the set of object states considered to be possible resultant states after applying the action with postcondition  $\alpha$  at the object state  $w$ . If for a certain  $w, \alpha$ , it happens that  $\mathcal{R}(w, \alpha) = \emptyset$ , this is taken to mean that  $\alpha$  is not applicable at  $w$ .

### 2.2 Extensions of the Object Language

Despite the many different ways in which the principle of minimal change has been encoded [McCarthy, 1980; Winslett, 1988; Katsuno and Mendelzon, 1992; Doherty, 1994;

<sup>2</sup>By a language, we intend all well formed formulae of that language.

Sandewall, 1996], a feature that is common to all these approaches is the existence of an ordering  $\leq$  on states used to determine which inferences are drawn about the effects of actions. In some of these approaches [Winslett, 1988; Katsuno and Mendelzon, 1992; Sandewall, 1996] the ordering  $\leq$  is defined over the set  $\mathcal{M}_{\mathcal{L}}$  of object states. For example, according to the Possible Models Approach (PMA), the ordering  $\leq_w$  associated with an (initial) state  $w$  is defined as follows: for any  $r, r' \in \mathcal{M}_{\mathcal{L}}$ ,  $r \leq_w r'$  if and only if  $\text{Diff}(w, r) \subseteq \text{Diff}(w, r')$ .<sup>3</sup> There are however many theories of action for which the ordering  $\leq$  is defined, not over the set of object states, but rather over an extended set of *meta-states*. Consider, for example, a theory of action based on circumscription [McCarthy, 1980]. Circumscription’s minimisation policy induces an ordering  $\leq$  that is defined over a set of meta-states  $\mathcal{M}_{\mathcal{L}'}$ , generated from the set of object states  $\mathcal{M}_{\mathcal{L}}$  with the addition of the *abnormality predicate*  $Ab$ . More precisely, if the object language has  $n$  fluents, and  $m$  actions, there will be  $2^n$  object states in  $\mathcal{M}_{\mathcal{L}}$ ; with the addition of the abnormality predicate  $Ab$ , each object state  $w$  “splits” into  $2^{n \times m}$  meta-states, all of which agree with  $w$  on the truth value of the  $n$  (object) fluents, and differ only on the value of the abnormality predicate  $Ab$  for each pair of (object) fluent and action. Thus there will be a total of  $2^{n+(n \times m)}$  meta-states over which the ordering  $\leq$  is defined.

As we prove later in this paper, moving the minimisation policy from object states to meta-states results in significant gains in the range of applicability of minimal change approaches. Given the major role of meta-states in our study, in the rest of this section we introduce some further notation and formally define the concepts related to meta-states.

A propositional language  $\mathcal{L}'$  is called an *extension* of  $\mathcal{L}$  if and only if firstly,  $\mathcal{L}'$  is *finitary* and secondly, the propositional variables of  $\mathcal{L}$  are included in  $\mathcal{L}'$ . If  $\mathcal{L}'$  is an extension of  $\mathcal{L}$  we shall refer to the additional propositional variables of  $\mathcal{L}'$  (i.e., those that do not appear in  $\mathcal{L}$ ) as *control variables* or *control fluents*.<sup>4</sup> We shall say that  $\mathcal{L}'$  is a  $k$ -extension of  $\mathcal{L}$ , for a natural number  $k \in \mathbb{N}$ , if and only if  $\mathcal{L}'$  is an extension of  $\mathcal{L}$  and it contains precisely  $k$  control fluents. Clearly, any 0-extension of  $\mathcal{L}$  is identical with  $\mathcal{L}$ .

For an extension  $\mathcal{L}'$  of  $\mathcal{L}$ , any maximally consistent set of literals of  $\mathcal{L}'$  is called a *meta-state*. For a set of sentences  $G$  of  $\mathcal{L}'$ , we define the *restriction of  $G$  to  $\mathcal{L}$* , denoted  $G/\mathcal{L}$ , to be the set  $G \cap \mathcal{L}$ . Finally, we define the restriction to  $\mathcal{L}$  of a collection  $V$  of *sets* of sentences of  $\mathcal{L}'$ , denoted  $V/\mathcal{L}$ , to be the set consisting of the restriction to  $\mathcal{L}$  of the elements of  $V$ ; in symbols,  $V/\mathcal{L}' = \{G/\mathcal{L} : G \in V\}$ .

### 2.3 Preferential Models

Having formally defined meta-states it remains to introduce a general model that encodes the concept of minimisation over meta-states.

**Definition 2.2** A preferential structure for  $\mathcal{L}$  is a triple  $\mathcal{U} = \langle \mathcal{L}', \mathcal{S}', \theta \rangle$  where:

<sup>3</sup>For any two states  $w, z$ ,  $\text{Diff}(w, z)$  denotes the symmetric difference of  $w$  and  $z$ .

<sup>4</sup>Like the abnormality predicate, control fluents are meant to be variables guiding the minimisation policy.

- $\mathcal{L}'$  is an extension of  $\mathcal{L}$ .
- $\mathcal{S}'$  is a nonempty collection of maximally consistent sets of literals of  $\mathcal{L}'$ ; we shall call the elements of  $\mathcal{S}'$  valid meta-states.
- $\theta$  is a function mapping each object state  $w \in \mathcal{M}_{\mathcal{L}}$  to a (partial) preorder over  $\mathcal{M}_{\mathcal{L}'}$  (the set of all maximally consistent sets of literals of  $\mathcal{L}'$ ); we shall denote the preorder assigned to  $w$ , by  $\leq_w$ .<sup>5</sup>

As mentioned earlier, a preferential structure  $\mathcal{U} = \langle \mathcal{L}', \mathcal{S}', \theta \rangle$  is meant to be the basis for encoding formally (and in a quite abstract manner) the concept of minimal change. More precisely, let  $w \in \mathcal{M}_{\mathcal{L}}$  be any object state. The preorder  $\leq_w$  associated with  $w$  represents the *comparative similarity* of meta-states to  $w$ . Using  $\leq_w$  (and the principle of minimal change), one can then determine the states  $\mathcal{R}(w, \alpha)$  that can possibly result from the application of an action  $\alpha$  to  $w$  by means of the condition (M) given below:

$$(M) \quad \mathcal{R}(w, \alpha) = (\min([\alpha]_{\mathcal{L}'}, <_w) \cap \mathcal{S}') / \mathcal{L}.$$

In the above condition,  $[\alpha]_{\mathcal{L}'}$  denotes the set of meta-states consistent with the sentence  $\alpha$  and  $\min([\alpha]_{\mathcal{L}'}, <_w)$  is the set of such meta-states that are minimal (“most preferred”) with respect to  $<_w$ .

The intuition behind condition (M) should be clear. Essentially, we select those meta-states consistent with formula  $\alpha$  (representing the postcondition of an action) that are minimal under the ordering  $<_w$ , filter out the valid ones and then restrict these to the language of the dynamic system under consideration.

We shall say that a preferential structure  $\mathcal{U} = \langle \mathcal{L}', \mathcal{S}', \theta \rangle$  is a *preferential model* for the dynamic system  $W = \langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$  if its result function can be reproduced from  $\mathcal{U}$  by means of condition (M); more precisely, if and only if for all  $w \in \mathcal{S}$  and  $\alpha \in \mathcal{A}$ , condition (M) is satisfied. If a dynamic system  $W$  has a preferential model  $M$ , we shall say that  $W$  is *minimisable*; moreover, if there are precisely  $k$  control fluents in  $\mathcal{L}'$ , we shall say that  $W$  is *k-minimisable* or that it has a preferential model with *degree k*. Clearly, if a dynamic system  $W$  has a preferential model with degree  $k$  for some  $k \in \mathbb{N}$  it also has a preferential model with degree  $m$  for any  $m \geq k$ .

### 3 Minimisable Dynamic Systems

Our aim in this section is to provide a characterisation of the class of minimisable dynamic systems, in terms of conditions imposed on the result function  $\mathcal{R}$ .

Let  $W = \langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$  be a dynamic system,  $w$  a state in  $\mathcal{S}$ , and  $\alpha, \alpha'$  any two actions in  $\mathcal{A}$ . Consider the conditions (P1) – (P3) below:

- (P1) If  $w \in \mathcal{S}$  and  $r \in \mathcal{R}(w, \alpha)$ , then  $r \vdash \alpha$
- (P2) If  $w \in \mathcal{S}$  and  $\vdash \alpha \leftrightarrow \alpha'$ , then  $\mathcal{R}(w, \alpha) = \mathcal{R}(w, \alpha')$
- (P3) If  $\alpha \vdash \alpha'$ ,  $w \in \mathcal{S}$ ,  $r \in \mathcal{R}(w, \alpha')$  and  $r \in [\alpha]$ , then  $r \in \mathcal{R}(w, \alpha)$

<sup>5</sup>We shall also use  $<_w$  to refer to the strict (non-reflexive) part of  $\leq_w$ .

These conditions can be interpreted quite simply. Condition (P1) says that the postconditions of an action (i.e.,  $\alpha$ ) should be true at all possible resultant states. Condition (P2) is an *irrelevance of syntax* condition stating that actions having logically equivalent postconditions should predict the same resultant states. Condition (P3) states that if a state  $r$  is chosen as a possible outcome of an action  $\alpha'$ , then  $r$  should also be chosen as a possible outcome of any action  $\alpha$  which is stronger than  $\alpha'$  and consistent with  $r$ . (NB:  $\alpha \vdash \alpha'$  implies  $[\alpha] \subseteq [\alpha']$ ). This last condition is similar to the choice theoretic condition known as  $(\alpha)$  in the literature [Sen, 1977]. These three, simply stated conditions suffice to exactly characterise the class of minimisable dynamic systems.

**Theorem 3.1** *A dynamic system is minimisable if and only if it satisfies the conditions (P1) – (P3).*

Theorem 3.1 is the central result of this article. What is perhaps surprising about this theorem is that it manages to provide a characterisation of minimality defined over *meta-states* via conditions on the result function, which operates on *object states*.

The proof of Theorem 3.1 is omitted due to space limitations. The most interesting part of the proof is a construction that, given a dynamic system  $W$  with  $n$  fluents satisfying (P1) – (P3), generates a preferential model for  $W$  with degree  $2^n$ . An immediate corollary of this is that, if a dynamic system is at all minimisable, then it is minimisable with degree  $2^n$ . We shall call the *smallest* number  $k$  for which a minimisable dynamic system  $W$  has a preferential model with degree  $k$ , the *minimality rank* of  $W$ . As already mentioned, the corollary below follows directly from the proof of Theorem 3.1.

**Corollary 3.1** *The minimality rank of a minimisable dynamic system with  $n$  fluents is no greater than  $2^n$ .*

Having provided a general characterisation of minimisable dynamic systems by means of (P1) – (P3), we shall now turn to special cases. More precisely, we shall impose certain constraints on preferential models and examine their implications on minimisable dynamic systems via condition (M).

The first such constraint is *totality* on the preorders of a preferential model. More precisely, we shall say that a preferential model  $\mathcal{U} = \langle \mathcal{L}', \mathcal{S}', \theta \rangle$  is *linear* if and only if all preorders in  $\theta$  are *total* (sometimes referred to as *connected*). We shall say that a dynamic system  $W = \langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$  is *strictly minimisable* if and only if  $W$  has a linear preferential model. In the presence of (P1) – (P3), condition (P4) below characterises precisely the class of strictly minimisable dynamic systems. The term  $\mathcal{R}(w, \mathcal{A}')$  in (P4), where  $\mathcal{A}'$  is a subset of  $\mathcal{A}$ , is used as an abbreviation for  $\bigcup_{\alpha' \in \mathcal{A}'} \mathcal{R}(w, \alpha')$ .

- (P4) For any nonempty  $\mathcal{A}' \subseteq \mathcal{A}$  such that  $\mathcal{R}(w, \alpha') \neq \emptyset$  for all  $\alpha' \in \mathcal{A}'$ , there exists a nonempty subset  $\mathcal{B}$  of  $\mathcal{R}(w, \mathcal{A}')$ , such that  $\mathcal{R}(w, \alpha) = [\alpha] \cap \mathcal{B}$ , for all  $\alpha \in \mathcal{A}$  such that  $[\alpha] \subseteq [\mathcal{A}']$  and  $[\alpha] \cap \mathcal{B} \neq \emptyset$ .

Condition (P4) essentially says that, under certain conditions, a collection of states  $V$  contains a subset  $\mathcal{B}$  of “best” elements. Consequently, whenever an action  $\alpha$  is such that all  $\alpha$ -states are contained in  $V$ , and moreover, among the  $\alpha$ -states there are some of the “best” elements of  $V$ , then any state resulting from  $\alpha$  is among those “best”  $\alpha$ -states (i.e.,  $\mathcal{R}(w, \alpha) =$

$[\alpha] \cap \mathcal{B}$ ). Notice that (P4) collapses to the (much more familiar) condition (P4') below, whenever  $\mathcal{R}(w, \alpha)$  is defined for all pairs of states  $w$ , and sentences  $\alpha$ . In the principal case however where  $\mathcal{R}$  is defined over a proper subset of  $\mathcal{M}_{\mathcal{L}} \times \mathcal{L}$ , (P4') is strictly weaker than (P4).

(P4') If  $\mathcal{R}(w, \alpha), \mathcal{R}(w, \alpha') \neq \emptyset, \mathcal{R}(w, \alpha) \subseteq [\alpha']$ , and  $\mathcal{R}(w, \alpha') \subseteq [\alpha]$ , then  $\mathcal{R}(w, \alpha) = \mathcal{R}(w, \alpha')$ .

This is essentially the (U6) postulate of Katsuno and Mendelzon [1992]. It is also found as property (3.13) in Gärdenfors [1988, p. 57].

**Theorem 3.2** *A dynamic system is strictly minimisable if and only if it satisfies the conditions (P1) – (P4).*

A direct consequence of the (only-if part of the) above proof is the following corollary.

**Corollary 3.2** *The minimality rank of a strictly minimisable dynamic system is no greater than 1.*

Corollary 3.2 shows that by imposing totality on the pre-orders of the preferential model, we get very close to zero-minimisable dynamic systems. Very close indeed, but not quite there. In this paper we do not provide a complete characterisation of zero-minimisable dynamic systems as it is not central to our aims here; we do however provide some preliminary results in this direction. More precisely, consider the conditions (P5) – (P7) below (we implicitly assume that  $w \in \mathcal{S}$  in each case):

(P5) If  $([\alpha] - \mathcal{S}) \cup \mathcal{R}(w, \alpha) \subseteq [\alpha']$ , then  $[\alpha] \cap \mathcal{R}(w, \alpha') \subseteq \mathcal{R}(w, \alpha)$ .

(P6) If  $[\alpha] \subseteq \mathcal{S}$ , then  $\mathcal{R}(w, \alpha) \neq \emptyset$ .

(P7)  $\mathcal{R}(w, \alpha) \cap \mathcal{R}(w, \alpha') \subseteq \mathcal{R}(w, \alpha \vee \alpha')$

Condition (P5) says that if all non-valid  $\alpha$ -states together with those  $\alpha$ -states “chosen” by the result function are compatible with another action’s postconditions ( $\alpha'$ ), then any  $\alpha$ -states chosen when performing  $\alpha'$  should also be chosen when performing  $\alpha$ . Notice that this condition implies condition (P3) above. (P6) states that the result function must return at least one possible resultant state if all states satisfying the postconditions of the action are valid. (P7) says that if a state  $r$  is chosen as a possible next state when either  $\alpha$  or  $\alpha'$  is performed, then  $r$  should also be chosen when the action (with postcondition)  $\alpha \vee \alpha'$  is performed.

**Theorem 3.3** *Every zero-minimisable dynamic system satisfies the conditions (P1) – (P3), (P5) – (P7).*

The converse of Theorem 3.3 is not true in general. However, for a restricted class of dynamic systems, which we call *dense*, the conditions (P1) – (P3) and (P5) – (P7) suffice to characterise zero-minimisability. More precisely, we shall say that a dynamic system is *dense* if and only if every sentence of the object language  $\mathcal{L}$  corresponds to an action (i.e.,  $\mathcal{A} = \mathcal{L}$ ).

**Theorem 3.4** *If a dense dynamic system satisfies (P1) – (P3), (P5) – (P7), then it is zero-minimisable.*

We conclude this section with a brief comment on previous frameworks that encode the concept of minimal change. Perhaps the framework most closely related to our own comes from the area of *theory change* and it is the one developed by Katsuno and Mendelzon [1992] for modelling *belief update*. We shall leave a detailed comparison between the conditions presented herein and the postulates proposed by Katsuno and Mendelzon (known as the *KM postulates*) for future work. Here we simply note that the main difference between the two is that the KM postulates are designed to apply *only* to dense, zero-minimisable, dynamic systems.

## 4 Causality and Minimal Change

Recall that one of the main motivations for this work was the desire to formally evaluate claims about the strength of causal theories of action over ones based on the notion of minimal change. In this section we use the foregoing results to analyse two of the most prominent causal theories of action, the first developed by McCain and Turner [1995], and the second by Thielscher [1997].

### 4.1 McCain and Turner

McCain and Turner [1995] have developed a theory of action that represents causality explicitly. In their framework they introduce a causal connective  $\Rightarrow$  where  $\phi \Rightarrow \psi$  can be read as “ $\phi$  causes  $\psi$ ” and referred to as a *causal rule*. Here  $\phi$  and  $\psi$  are propositional sentences that do not contain  $\Rightarrow$  (i.e.,  $\Rightarrow$  cannot be nested). They then introduce the notion of *causal closure*  $C_{\mathcal{D}}(\Gamma)$  for a set of sentences  $\Gamma$  with respect to a set of causal rules  $\mathcal{D}$  as the smallest set closed under classical deduction that includes  $\Gamma$  and applies causal rules in the direction of the “arrow” (i.e., no contrapositive—the interested reader is referred to the citation above for the full details). The notation  $\vdash_{\mathcal{D}}$  refers to the corresponding (causal) consequence relation:  $\Gamma \vdash_{\mathcal{D}} \gamma$  if and only if  $\gamma \in C_{\mathcal{D}}(\Gamma)$ . The result function can be defined using the following fixed-point equation to be found in McCain and Turner [1995].

(MT)  $r \in \mathcal{R}_{\mathcal{D}}^{MT}(w, \alpha)$  iff  $r = \{\rho \in \mathcal{Z}_{\mathcal{L}} : (w \cap r) \cup \{\alpha\} \vdash_{\mathcal{D}} \rho\}$

We can now establish the following results.

**Theorem 4.1** *For any set of causal rules  $\mathcal{D}$ , the result function  $\mathcal{R}_{\mathcal{D}}^{MT}(w, \alpha)$  defined by means of (MT), satisfies the conditions (P1) – (P3).*

From Theorem 4.1 it follows that the theory of action developed by McCain and Turner is applicable *only* to minimisable dynamic systems. This is a most curious result for it shows that the conclusions drawn with the aid of causality (as encoded by McCain and Turner) can be reproduced by an appropriately defined minimisation policy. It is also especially interesting in light of recent results reported by Peppas *et al.* [1999] who show that for McCain and Turner’s approach it is in general not possible to construct an ordering over object states such that the minimal object states satisfying the postconditions of an action are those predicted by the McCain and Turner fixed-point equation. This tells us that this approach is not zero-minimisable. We have, however,

just shown that this approach is minimisable in a more general sense (i.e., if one considers meta-states) so this system has a minimality rank greater than zero.

## 4.2 Thielscher

We shall not describe Thielscher’s [1997] approach in depth here. However, the underlying principle is to consider trajectories of state-effect pairs each of which is the result of applying a causal law at a previous state-effect pair (starting with the initial state and action postcondition). The resultant states are those at the end of a trajectory where causal laws no longer apply. We note that Thielscher’s system does not satisfy postulates (P1) – (P3) and hence is not minimisable. In particular, Thielscher’s result function, which we denote by  $\mathcal{R}^T$  violates condition (P1); the postconditions of the action do not necessarily have to hold after applying the action. What would be of some interest however, is to generate from  $\mathcal{R}^T$  a new result function  $\mathcal{R}'$  that chooses among the resultant states selected by  $\mathcal{R}^T$  the ones that satisfy the postconditions of the occurring action; in symbols,  $\mathcal{R}' = [\alpha] \cap \mathcal{R}^T$ . One can then examine whether the new function  $\mathcal{R}'$  satisfies the conditions (P2) and (P3). We leave this for future investigation.

## 5 Discussion

Let us take a step back and examine what we have accomplished thus far. The main result reported herein is a characterisation of the class of dynamic systems amenable to minimisation, in terms of conditions on the result function. Existing causal theories of actions can then be assessed against these conditions to determine the added value (if any) of explicit representations of causality. This is clearly a significant step towards “demystifying” the (comparative) strengths and weaknesses of the notions of causality and minimal change in reasoning about action. Admittedly though in this paper we have not given the whole story (especially as far as “demystifying” causality is concerned). What is missing is a generic model of the use of the concept of causality in reasoning about action (in the same way that preferential structures are such a model for the concept of minimality), based on which a general, formal comparison between causality and minimality can be made. Until such a generic model is available, the best that can be done is evaluations of *specific* causal theories of action (such as the ones by McCain and Turner [1995] or Thielscher [1997]) against the conditions (P1) – (P3). There is of course still a lot of value in such assessments; showing for example that all existing causal approaches satisfy these conditions would be strong evidence in support of the claim that minimality subsumes causality (at least as far as the range of applicability is concerned). Showing, on the other hand, that some causal approach violates one of the conditions (P1) – (P3), would prove that causality is essential in reasoning about action since it covers domains that are “unreachable” by minimal change approaches. Similar (although weaker) conclusions can be drawn from the satisfaction or violation of (P4) and (P5) – (P7). We have already witnessed that Thielscher’s [1997] approach does not satisfy condition (P1). Thus some causal approaches do indeed go beyond what is possible with minimal change. More work needs to be done

here to properly classify causal approaches both in regard to each other (with respect to the different causal notions they employ) and also with regard to the Principle of Minimal Change.

It should be noted, that when assessing a theory of action, apart from its range of applicability, a second criterion that is equally important is the *conciseness* of its representations (a solution to the *frame problem* ought to be both *correct* and *concise*). In this article, conciseness has been left out of the picture altogether. What we have mainly done herein was to axiomatically characterise certain classes of dynamic systems whose result function  $\mathcal{R}$  can be reproduced in terms of preorders  $\leq_w$  on states. No consideration was given as to whether  $\leq_w$  can be represented concisely or not. As far as our results are concerned, describing  $\leq_w$  could be as “expensive” as listing the frame axioms corresponding to  $\mathcal{R}$ , which of course defeats the whole purpose of using minimality. A much more useful result (for practical purposes) would be one that characterises the class of what we might call *concisely* minimisable dynamic systems; that is, dynamic systems whose result function  $\mathcal{R}$  can be reproduced by preorders  $\leq_w$ , which in turn can be represented concisely.<sup>6</sup>

Similar considerations apply to causality. Characterising the class of dynamic systems for which causality (in one form or another) can duplicate the result function, although an interesting theoretical result, would not fully address the issue of the applicability of causality in reasoning about action. Further work would be required to identify the domains that are amenable to concise causal descriptions.

Notice, also, that in cases where the range of applicability does not differentiate between causal and minimality-based approaches, conciseness considerations may well favour one over the other. More precisely, if the class of domains at focus is within the range of applicability of both causal and minimal change approaches, the determining factor in choosing between the two could be the “information cost” associated with the usage of each approach.

Notice that, while the concept of minimality has typically been used in the literature to deal with the frame and ramification problems, the nature of condition (M) is such that it requires minimality to deal with the *qualification problem* as well. Indeed, via condition (M), the preorders of a preferential model are used, not only to determine the set of states  $\mathcal{R}(w, \alpha)$  that result from the application of an action  $\alpha$  at an initial state  $w$  (frame and ramification problems), but also to determine whether  $\alpha$  is at all *applicable* at  $w$  (qualification problem). The latter is decided based on whether  $\mathcal{R}(w, \alpha)$  is the empty set or not (cf. McCain and Turner [1995] who claim this is a *derived qualification*). One could argue that such an additional burden is perhaps too much from minimality to carry (or that it is even counter-intuitive), and maybe, if liberated from it, minimality could be used in many more situations. More formally, consider the condition (M') below:

<sup>6</sup>It should be noted that all existing approaches that are based on the concept of minimal change are indeed of that nature; that is, their preorders on states have a concise description, typically in the form of a second-order axiom (sometimes coupled with limited domain-specific information).

(M') If  $\mathcal{R}(w, \alpha) \neq \emptyset$ , then  $\mathcal{R}(w, \alpha) = (\min([\alpha]_{\mathcal{L}'}, \leq_w) \cap \mathcal{S}') / \mathcal{L}$ .

Clearly (M') is weaker than (M). In fact, it is exactly the weakening of (M) that is needed to disengage minimality from the qualification problem. It would be a worthwhile exercise to reproduce the results of this article, having replaced (M) with (M'). The class of minimisable systems could be larger, and moreover, we conjecture that under (M'), strictly minimisable systems are a proper subclass of zero-minimisable dynamic systems.

We conclude this section with a remark on minimality ranks. Preliminary considerations suggest that there is a close relation between the minimality rank of a dynamic system on the one hand, and its ontological properties on the other. For example, domains where actions have no ramifications, tend to have lower minimality ranks than domains where ramifications do appear. If this connection can be generalised and formally proved, the minimality rank of a domain could serve as a precise measure of its complexity. More work however needs to be done in this direction.

## 6 Conclusions

We have developed a formal framework within which we were able to formulate necessary and sufficient conditions under which a dynamic system is minimisable; that is, its result function can be reproduced by an appropriately defined minimisation policy. What is particularly pleasing about these conditions is that they are few in number and relatively easy and intuitive to understand. Our original motivation in this study was to answer the question as to whether recently proposed theories of action involving explicit representations of causality are capable of forms of reasoning not possible via minimal change. We have seen that this is not the case with some approaches (McCain and Turner [1995]) but in others (Thielscher [1997]) the causal reasoning cannot be characterised by the Principle of Minimal Change. Perhaps of wider significance is the fact that the results reported here clearly indicate the range of applicability of the Principle of Minimal Change; one simply needs to verify three properties (viz. (P1) – (P3)).

This work opens up many interesting avenues for future work, various of which were mentioned in the discussion. One of the more pressing is to consider a variety of theories of action, particularly causal ones, and verify which properties they satisfy. This task is well under way and reserved for a much lengthier work. Also of much interest would be to further categorise levels of minimisability and what distinguishes them together with the various theories of action at those levels of minimisability.

## Acknowledgments

The authors would like to thank three anonymous referees for their insightful and generous comments. They would also like to thank Norman Foo, Abhaya Nayak, Mikhail Prokopenko, members of the Cognitive Robotics Group at the University of Toronto, and attendees of the Seminar in Applications of Logic at the City University of New York for enlightening discussions.

## References

- [Doherty, 1994] P. Doherty. Reasoning about Action and Change using Occlusion. In *Proceedings of the Eleventh European Conference on Artificial Intelligence*, pp. 401–405, 1994.
- [Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux*. The MIT Press, Cambridge, MA, 1988.
- [Katsuno and Mendelzon, 1992] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In P. Gärdenfors, ed., *Belief Revision*, pp. 183–203, 1992.
- [Krautz, 1986] A. Krautz. The logic of persistence. in *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 401–405, 1986.
- [Lifschitz, 1990] V. Lifschitz. Frames in the space of situations, *Artificial Intelligence*, **46**:365–376, 1990.
- [Lin, 1995] F. Lin. Embracing causality in specifying the indeterminate effects of actions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 2001–2007, 1995.
- [McCain and Turner, 1995] N. McCain and H. Turner. A causal theory of ramifications and qualifications. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1978–1984. Montreal, 1995.
- [McCain and Turner, 1997] N. McCain and H. Turner. Causal theories of action and change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pages 460–465, 1997.
- [McCarthy, 1980] J. McCarthy. Circumscription—A form of nonmonotonic reasoning. *Artificial Intelligence*, **13**:27–39, 1980.
- [Peppas et al., 1999] P. Peppas, M. Pagnucco, M. Prokopenko, N. Foo and A. Nayak. Preferential semantics for causal system. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 119–123, 1999.
- [Sandewall, 1989] E. Sandewall. Filter preferential entailment for the logic of action in almost continuous worlds. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 1989.
- [Sandewall, 1994] E. Sandewall. *Features and Fluents. The Representation of Knowledge about Dynamical Systems. Vol. 1*. Oxford University Press, 1994.
- [Sandewall, 1996] E. Sandewall. Assessments of ramification methods that use static domain constraints. In *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*. 1996.
- [Sen, 1977] A. Sen, Social choice theory: A re-examination, *Econometrica*, **45**:53–89, 1977.
- [Shoham, 1988] Y. Shoham. *Reasoning About Change*. MIT Press, Cambridge, Massachusetts, 1988.
- [Thielscher, 1997] M. Thielscher. Ramification and causality. *Artificial Intelligence*, **89**:317–364, 1997.
- [Winslett, 1988] M. Winslett. Reasoning about actions using a possible models approach. In *Proceedings of the Seventh National Artificial Intelligence Conference*, 1988.