Intensionality and Circumscription

Dr. G. White
Department of Computer Science
Queen Mary, University of London
graham@dcs.qmul.ac.uk
http://www.dcs.qmul.ac.uk/~graham

1 Introduction

This paper examines the semantics of a group of logical techniques, used by the Artificial Intelligence community, and which are fundamental for circumscription-based approaches to the situation calculus. At first sight, this calculus appears to be an operationalised version of David Lewis' account of causal reasoning: there are items – here called 'situations' – which are analogous to Lewis' possible worlds, and one finds the situation which is the result of an action by minimising difference: that is, if an action is performed in a situation s, then the situation resulting from the action is that situation which, among all of those in which the effects of the action hold, is minimally different from s.

I will argue, however, that appearances deceive: that, whereas Lewis' theory is extensional (in the sense that equal possible worlds can always be substituted for each other without harm), the situation calculus is strongly intensional. Consequently, the underlying ontologies of the two theories are strikingly different. The culprit here is the way in which the situation calculus measures difference: whereas Lewis' measure is required to be extensional (that is, to depend only on the identities of the worlds involved), the measure used in the situation calculus depends strongly on the way in which the worlds are described: we can have two languages which are logically equivalent, but which, nevertheless, yield distinct difference measures, and which, consequently, give different results for the same problem when used in the situation calculus.

The situation calculus, then, is not logically innocuous: if it can be given any principled semantics, then that semantics must involve an extremely intensional ontology.

2 The Situation Calculus and an Example

2.1 The Situation Calculus

I shall describe the situation calculus rather generically; the phenomena that I am describing apply to a great number of versions of the situation calculus, and consequently a rather more abstract approach has the benefit of clarity.

Situations are described by *fluents*: a fluent is a linguistic item – typically a propositional atom – that can be true or false in a given situation. So there are two sides to fluents: on the one hand, since they can be assigned truth values, they have a propositional role. On the other hand, they are tied to a particular language, and, specifically to a particular vocabulary, in which they figure as atoms.

For brevity, I shall be making my arguments informally, in terms of finding situations which minimise change, and, correspondingly, in terms of the partial order on the sets of models given by inclusion of the sets of changed fluents: the arguments are easy enough to formulate in terms of circumscription, in the style of (Shanahan, 1997, pp. 49ff) or of (Lifschitz, 1994, p. 329).

2.2 The Example

Consider the following scenario:

> Two balls are linked by a string of length 1. Initially (at $t = 0$) they are lying next to one another on a table, with the string slack. At $t = 1$, one ball is raised to height 2; this tightens the string and raises the other ball to height 1.

Now let us consider formulations of this scenario in several different vocabularies. In all cases the balls will be b_1 and b_2, and the positions will simply be numbered 0, 1, 2, 3, starting from the table and working upwards, so that the two balls start off in position 0.

With some trivial work, we can show that all of our vocabularies are logically equivalent: any assertion made in one of them can be expressed without loss in any of the others. The constraints on the fluents are, likewise, equally easy to formulate in any vocabulary: they simply say that, if one of the balls is at position i, then the other ball must either be at positions $i-1$ (if this is positive) or $i+1$, with the string taut, or at i, with the string slack.

Because of the constraints, ball 2 must either end up at positions 1 or 3, with the string taut, or at 2, with the string slack. So there are basically three situations to consider; I will call them m_1, m_2, and m_3 in what follows. I tabulate the order relations induced by a selection of vocabularies in Table 1

3 Formal Results on Interpretation

As we can see from the tables, we appear to be able to obtain a wide range of partial orders merely by change of vocabulary. As the results of this section show, this variability is quite typical.

We first need the notion of an interpretation of one language in another. Notice that, even though our languages in the preceding section was ostensibly first-order, its variables only take on a finite range of values: we can, thus, for simplicity, assume that our languages are propositional and have countable sets of atoms. They will have the classical connectives, $\wedge$, $\vee$, $\rightarrow$, and $\neg$.

Definition 1. *Let $\mathfrak{L}$ and $\mathfrak{L}'$ be two propositional languages. An* interpretation*, α, of $\mathfrak{L}$ in $\mathfrak{L}'$ is a map from the atoms of $\mathfrak{L}$ to the propositions of $\mathfrak{L}'$; it may be extended to all of $\mathfrak{L}$ inductively.*

Suppose we have theories, τ and τ', formulated respectively in the languages $\mathfrak{L}$ and $\mathfrak{L}'$. We will identify theories with deductively closed sets of sentences.

Definition 2. *An interpretation α is* sound *(with respect to τ and τ') if, whenever T is a theorem of τ, $\alpha(T)$ is a theory of τ'; it is* conservative *(again, with respect to τ and τ') if, given a proposition p of $\mathfrak{L}$, if $\alpha(p)$ is a theorem of τ', then p is a theorem of τ. α is* essentially surjective *(with respect to τ') if, for all $q \in \mathfrak{L}'$, there is a $p \in \mathfrak{L}$ such that $\tau \vdash \alpha(p) \leftrightarrow q$.*

We are concerned with these theories mainly for the sake of their effect on models:

Definition 3. *A* valuation *of a language $\mathfrak{L}$ is an assignment $\nu(\cdot)$ of truth values to its atoms; this can be extended to the propositions of $\mathfrak{L}$ in the usual way. Such an assignment is a* model *of τ if all theorems of τ are assigned* $\mathfrak{t}$.

Suppose, then, that we have a sound interpretation α of $\mathfrak{L}$ in $\mathfrak{L}'$; if ν' is a model of τ', then, by soundness, $\nu' \circ \alpha$, the composition of α with ν, is a model of τ. We thus have a mapping, α^*, from models of τ' to models of τ.

The proofs of the following results are quite straightforward.

Proposition 1. *If α is sound, then α is conservative iff α^* is surjective.*

Proposition 2. *If α is sound, then α is essentially subjective iff $\alpha*$ is injective.*

3.1 Applications

Our applications of these results will be in very simple cases; we will typically start with two different languages, $\mathfrak{L}$ and $\mathfrak{L}'$, and an interpretation α of $\mathfrak{L}$ in $\mathfrak{L}'$. We will also start with a theory in one language or the other, and will define a theory in the other to correspond to it, either by starting with a theory $\tau \subset \mathfrak{L}$ and defining τ' to be the deductive closure of $\alpha(\tau)$, or by starting with a theory $\tau' \subset \mathfrak{L}'$ and defining τ to be $\alpha^{-1}(\tau')$. In either case, soundness is a triviality. We will also be working in situations where it is easy to establish both conservativity and essential surjectivity; so, in consequence, α^* will be a bijective correspondence on models.

We need, however, some more elaborate notation for theories. We are concerned with situations which evolve: for simplicity, our systems will only evolve from time 0 to time 1. However, this means that we have three theories to deal with: we will have a *base theory*, $\underline{\tau}$, consisting of the constraints which are valid atemporally, and we will also have two extensions of this, τ_0 and τ_1, which consist of $\underline{\tau}$ together with the constraints which are valid at times 0 and 1. We will have to verify, rather tediously, our conditions for the correspondences between models of all three theories; we can save some work because, if α is essentially surjective with respect to $\underline{\tau}$, it will be so with respect to τ_0 and τ_1.

We are interested in the $\preceq$ relation between models at time 1 (i.e. models of τ_1): in our situation, this is defined as follows. Suppose we have a language, theories in that language as above, and a model ν_0 of

<table>
<tr><th>Vocabulary</th><th>Fluents</th><th>Order</th></tr>
<tr><td>A</td><td>$\mathsf{at}(b_i, j)|_{i=1}^{2}\,{}_{j=0}^{3}$
taut</td><td>m_1 m_2 m_3</td></tr>
<tr><td>B</td><td>$\mathsf{above}(b_i, j)|_{i=1}^{2}\,{}_{j=0}^{3}$
taut</td><td>m_3
|
m_1 m_2</td></tr>
<tr><td>C</td><td>$\mathsf{above}(b_i, j)|_{i=1}^{2}\,{}_{j=0}^{3}$</td><td>$m_3$
|
m_2
|
m_1</td></tr>
<tr><td>D</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{above}(b_1, b_2)$, $\mathsf{above}(b_2, b_1)$</td><td>m_1 m_3
\ /
m_2</td></tr>
<tr><td>E</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{above}(b_2, b_1)$
taut</td><td>m_3
|
m_1
|
m_2</td></tr>
<tr><td>F</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{above}(b_1, b_2)$
taut</td><td>m_1
|
m_3
|
m_2</td></tr>
<tr><td>G</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{at}(b_2, 2)$, $\mathsf{at}(b_2, 3)$</td><td>m_2 m_3
\ /
m_1</td></tr>
<tr><td>H</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{at}(b_2, 1)$, $\mathsf{at}(b_2, 2)$</td><td>m_1 m_2
\ /
m_3</td></tr>
<tr><td>I</td><td>$\mathsf{above}(b_1, j)|_{j=0}^{2}$
$\mathsf{above}(b_2, 1)$
$\mathsf{at}(b_2, 2)$</td><td>m_2
|
m_3
|
m_1</td></tr>
</table>

Table 1: Vocabularies and Their Orders

τ_0. Then we can define a relation as follows:

Definition 4. *If ν is a model of τ_1, define the difference set of ν to be*

$$\delta_{\mathfrak{L},\nu_0}(\nu) \;=\; \{p \in \mathfrak{L} | p \text{ an atom}, \nu(p) \neq \nu_0(p)\}$$

We then define the order relation, $\preceq_{\mathfrak{L},\nu_0}$ by inclusion of difference sets:

$$\mu \preceq_{\mathfrak{L},\nu_0} \nu \;\Leftrightarrow\; \delta_{\mathfrak{L},\nu_0}(\mu) \subseteq \delta_{\mathfrak{L},\nu_0}(\nu)$$

Notice that we have carefully made explicit the dependence of $\preceq_{\mathfrak{L},\nu_0}$ on both ν_0 and on $\mathfrak{L}$.

Now we have the following

Proposition 3. *Suppose that we have a language $\mathfrak{L}$ together with a theory τ. Then, given a language $\mathfrak{L}'$ containing $\mathfrak{L}$, and a theory τ' such that the inclusion is sound, conservative and essentially surjective, we have*

$$\preceq_{\mathfrak{L}',\nu_0} \subseteq \preceq_{\mathfrak{L},\nu_0}$$

i.e. as we enlarge the language we restrict the accessibility relation on models.

Proof. Clear

□

Example Language B is such an extension of Language C. This example shows that the containment is, in general, strict.

This example prompts the question of how far we can go: whether, given any language, we can extend it so as to make the order relation discrete. It turns out that we can; however, we need another result first.

Proposition 4. *Suppose we have a language, $\mathfrak{L}$, and a theory τ. Suppose that we are given a model, ν_0, of τ, and an order relation $\preceq$ on the models of τ for which ν_0 is the unique least element. Then there is a language $\mathfrak{L}'$, a theory τ', and a sound, conservative and essentially surjective interpretation α of $\mathfrak{L}$ in $\mathfrak{L}'$ such that $\preceq_{\mathfrak{L}',\nu_0} = \preceq$.*

Proof. Before proving the proposition, we first standardise some notation. Let $\mathfrak{M}$ be the set of models of $\mathfrak{L}, \tau$. For each $\nu \in \mathfrak{M}$, take a propositional atom p_ν; let $\mathfrak{L}'$ be the language generated by these propositional atoms. Then define an interpretation α of $\mathfrak{L}$ in $\mathfrak{L}'$ by, firstly, defining, for each model ν of τ, a formula of $\mathfrak{L}'$

$$\vartheta_\nu \;=\; \left(\bigwedge_{\nu' \preceq \nu} p_{\nu'} \wedge \bigwedge_{\nu' \not\preceq \nu} \neg p_{\nu'} \right)$$

and then defining the interpretation α of an atom q of $\mathfrak{L}$:

$$\alpha(q) \;=\; \bigvee_{\nu(q)=\mathfrak{t}} \vartheta_\nu \qquad (1)$$

Let also τ' be given by the single proposition

$$\alpha(\top) \;=\; \bigvee_{\nu} \theta_\nu$$

We can then show that α and τ' give the desired interpretation. □

Given a language, and a desired partial order, this result gives us a *disjoint* language together with an interpretation which realises the partial order in terms of model similarity. The following result shows us what we can do if we want an *enlargement* of the original language: in this case, Proposition 3 shows us that all we can ask for is a subrelation of the original relation $\preceq_{\mathfrak{L},\nu_0}$. But we can achieve any subrelation:

Proposition 5. *Suppose we have a language, $\mathfrak{L}$, a theory τ, and a model ν_0 of $\mathfrak{L}$; we thus have a relation $\preceq_{\mathfrak{L},\nu_0}$ on models of τ. Then, given any reflexive transitive relation $\preceq$ contained in $\preceq_{\mathfrak{L},\nu_0}$, we can find a sound, complete and essentially surjective inclusion of languages $\mathfrak{L} \hookrightarrow \mathfrak{L}'$ such that $\preceq_{\mathfrak{L}',\nu_0} = \preceq$.*

Proof. Suppose given $\mathfrak{L}$, τ and ν_0, together with a desired relation $\preceq \; \subseteq \; \preceq_{\mathfrak{L},\nu_0}$. By Proposition 4, we can find a language $\mathfrak{L}'$ together with an interpretation α which realises $\preceq$. Now let $\mathfrak{L}''$ be the language whose set of atoms is the disjoint union of the sets of atoms of $\mathfrak{L}$ and $\mathfrak{L}'$, and expand τ with the definitions $p \leftrightarrow \alpha(p)$ for all atoms p of $\mathfrak{L}$; call this theory τ''. Then $(\mathfrak{L}'', \tau'')$ is the required extension of $(\mathfrak{L}, \tau)$:

$$\preceq_{\mathfrak{L}'',\nu_0} = \preceq_{\mathfrak{L},\nu_0} \cap \preceq,$$

and, since $\preceq \; \subseteq \; \preceq_{\mathfrak{L},\nu_0}$. we have the result. □

3.2 Interpretations of Our Vocabularies

We have still to justify our informal description of the mutual interpretations between our different vocabularies. First notice that all of the vocabularies are embedded in a single vocabulary (call it $\hat{\mathfrak{L}}$) which is generated by the union of all of their atoms. There is also a theory, $\hat{\tau}$, which expresses the constraints between these atoms; this theory will have precisely three models, corresponding to our three situations. Given one of our vocabularies, $\mathfrak{L}_i$, we can embed it in $\hat{\mathfrak{L}}$ in the obvious way; let τ_i be $\hat{\tau} \cap \mathfrak{L}_i$. The inclusion is clearly sound and complete. To show that it is essentially surjective, we have to show that each atom of $\hat{\mathfrak{L}}$ can be defined in $\mathfrak{L}_i$; and that can be checked case by case.

4 Inference Relations

What is interesting here is not so much the correctness, or otherwise, of the answers given by particular vocabularies as the fact that the various vocabularies give *different* answers.

We might think that this is nothing to worry about; that the differences are, to be sure, syntactic rather than model-theoretic, but that they would appear natural given a more syntactic approach to nonmonotonic reasoning, such as the inference relations of Kraus, Lehmann and Magidor (Gabbay, 1985; Makinson, 1989; Makinson, 1994; Kraus *et al.*, 1990). In this approach, we typically have two inference relations: a classical relation, $\vdash$, and a non-monotonic one, $\mid\!\sim$; the non-monotonic one is supposed to satisfy the rules in Table 2.

We need an alternative description of such inference relations: first, some notation.

Definition 5. *Suppose that $\mathfrak{L}$ is a propositional language, and that $\mathfrak{M}$ is its set of models. Then, for $p \in \mathfrak{L}$, let $\mu(p) = \{\nu \in \mathfrak{M} | \nu(p) = \mathfrak{t}\}$.*

Then we have

Proposition 6. *Given an inference relation $\mid\!\sim$ satisfying the rules in Table 2, there is exactly one map $\eta : \mathfrak{P}(\mathfrak{M}) \to \mathfrak{P}(\mathfrak{M})$ such that*

$$p \mid\!\sim q \text{ iff } \eta(\mu(p)) \subseteq \mu(q). \tag{2}$$

Conversely, any such η gives rise to exactly one inference relation $\mid\!\sim$.

Proof. Suppose given an inference relation $p \mid\!\sim q$; by Left Logical Equivalence, for any q, if $\mu(p) = \mu(\tilde{p})$, then $p \mid\!\sim q$ iff $\tilde{p} \mid\!\sim q$. Similarly, by Right Weakening, if $\mu(q) = \mu(\tilde{q})$, then, for any p, $p \mid\!\sim q$ iff $p \mid\!\sim \tilde{q}$. $\mid\!\sim$ may thus be regarded as a relation – $\mathcal{R}$, let us say – on $\mathfrak{M} \times \mathfrak{M}$. Now fix $A \subseteq \mathfrak{M}$ (i.e. $A \in \mathfrak{P}(\mathfrak{M})$). Right Weakening entails that, if $A\mathcal{R}B$ and if $B \subseteq \tilde{B}$, then $A\mathcal{R}\tilde{B}$. Similarly, if $A\mathcal{R}B$ and $A\mathcal{R}\tilde{B}$, the And rule tells us that $A\mathcal{R}B \cap \tilde{B}$. Because we have only a finite number of models, any descending chain $\{B_i\}$ of sets of models with $A\mathcal{R}B_i$ must terminate. Thus, for our fixed A, there is a unique minimal B_0 such that $A\mathcal{R}B_0$; thus, by Right Weakening, $A\mathcal{R}B$ iff $B_0 \subseteq B$. Let $\eta(A)$ be this B_0.

Conversely, given an $\eta : \mathfrak{P}(\mathfrak{M}) \to \mathfrak{P}(\mathfrak{M})$, we can, by (2), define a relation $\mid\!\sim$ on $\mathfrak{L} \times \mathfrak{L}$, and it is trivial to verify that it satisfies the conditions of Table 2. $\square$

Example 1. Given a partial order, $\preceq$, on $\mathfrak{M}$, we can define $\eta(A) = \{a \in A | \nexists \tilde{a} \in A. \tilde{a} \prec a\}$ and recover the usual definition of a non-monotonic inference relation induced by a model preference relation.

So now we have our argument. Suppose we have an entailment $A \mid\!\sim B$, formulated in one language $\mathfrak{L}$, and suppose also we have an interpretation α of $\mathfrak{L}$ in $\mathfrak{L}'$ which induces a bijection on models; then, by the above results, we can just as well formulate $\mid\!\sim$ in $\mathfrak{L}'$. Nonmonotonic inference relations may well be syntactically *formulated*, but, because of their logical properties, they are independent of the particular logical *language* that they are formulated in.

In fact, more than this invariance under *equivalent* vocabularies is true: it is also possible to establish that *any* interpretation (whether it is a bijection on models or not) induces mappings with good properties, in both directions, between nonmonotonic relations formulated in each language.

5 Comparisons

The reason why predictions are different in our various vocabularies seems to be this. Vocabulary plays a dual role in the situation calculus. On the one hand, it furnishes terms out of which the language of the theory is constructed. On the other hand, it plays a role in the measurement of change: changes are more or less according to whether their respective sets of changed fluents are included in each other, and fluents are items in the vocabulary. (Or, similarly, the vocabulary plays a role in circumscribing predicates: we circumscribe predicates by minimising the intersection of their extension with the set of fluents).

In the semantics of the situation calculus, then, we should be concerned, not just with the relation between the terms of its formal language and reality: that is adequately described by model theory, and we know that model theory behaves well under change of vocabulary. We should also be concerned with the auxiliary notions of the calculus: the closeness relations, or the way that the circumscription relations are constructed.

5.1 David Lewis

Contrast this with David Lewis' account of counterfactuals (Lewis, 1986). He defines the truth-conditions for counterfactual conditionals in terms of an accessibility relation on a set of possible worlds. His counterfactual connective, $\Box\!\!\rightarrow$, has the following semantics:

> $A \Box\!\!\rightarrow C$ is true at i iff some (accessible) $A \wedge C$-world is closer to i than any $A \wedge \neg C$-world, if

$\dfrac{\alpha \vdash \beta}{\alpha \mid\!\sim \beta}$	$\dfrac{\alpha \dashv\vdash \beta \quad \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma}$	$\dfrac{\alpha \mid\!\sim \beta \quad \vdash \beta \rightarrow \gamma}{\alpha \mid\!\sim \gamma}$	$\dfrac{\alpha \mid\!\sim \beta \quad \alpha \mid\!\sim \gamma}{\alpha \mid\!\sim \beta \wedge \gamma}$
Supraclassicality	Left Logical Equivalence	Right Weakening	And

Table 2: Rules for Non-Monotonic Inference Relations

> there are any (accessible) A-worlds. (Lewis, 1986, p. 10)

Closeness, here, is defined in terms of a triadic relation $j \preceq_i k$, which has certain formal properties; it plays a similar role in Lewis' theory to the fluent-based comparisons of the situation calculus.

The vital difference, though, is that it closeness is a relation between *possible worlds*; and worlds, for Lewis, are elements of reality.[1]

Our situation calculus comparisons cannot be interpreted, then, in terms of Lewis' closeness relations. Since descriptions of the same situations in our two vocabularies describe the same states of affairs – and how could they not, since the the vocabularies are inter-translatable without loss – they are descriptions of the same worlds; and any Lewis-style closeness relation would have to give the same results in both cases. The situation calculus relations are thus not relations between propositional worlds, but relations between propositional descriptions of those worlds.

6 Semantics and Identity

But why should this matter? There are two main reasons, theoretical and practical: they are closely intertwined, but somewhat distinct. We will start with the theoretical reasons.

Quine distinguishes between two levels of semantic comprehension. The first is that of merely responding appropriately to stimuli; this involves a child, for example,

> learning how much of what goes on about him counts as the mother, or as red, or as water. . . . Hello! more mama, more red, more water. (Quine, 1969, 7)

Later on, however, we learn to individuate the items that we encounter:

> To learn 'apple' it is not sufficient to learn how much of what goes on counts as an apple; we must learn how much counts as *an* apple, and how much as another. . . . [the child] can never fully master 'apple' in its individuative use, except as he gets on with the scheme of enduring and recurrent physical objects. (Quine, 1969, 8)

And thus

> Words like 'apple', and not words like 'mama' or 'water' or 'red', are the terms whose ontological involvement runs deep. (Quine, 1969, 8)

What is at issue here is that these two levels of understanding – the pre-identity stage, and the identity stage – are distinct, and that one may comprehend one without comprehending the other. So a piece of theory – the situation calculus, for example – may succeed in correlating its theoretical terms with the phenomena that it is trying to explain, and may seem to do so quite successfully; but, unless it can go further, and give well-motivated conditions of identity for the theoretical entities that it is trying to introduce, one may well feel that there is something lacking.

7 Excuses

We still need some further argument: we need to know that the problem with change of vocabulary is, indeed, a problem for our common-sense grasp of action and change. If it were not, then we might well argue that things were perfectly correct as they stood: that situations (as described in the situation calculus) could be given conditions of identity, but that they were more fine-grained than the conditions of identity for possible worlds. Indeed, Costello and MacCarthy give an account of counterfactual conditionals in which claims very like these are made. (Costello & McCarthy, 1999) So, if we had such a fine-grained account, situations in one vocabulary could maybe distinct from situations in another vocabulary. This is not so: but we need some argument to show it.

The first stage of the argument uses our metatheoretical results: we can achieve, by appropriate choice of vocabulary, *any* partial order on models that we desire. So, if we want to talk about the mathematical

[1] See (Lewis, 1983c, p. 39) "A world is a large possible individual; it has smaller possible individuals as parts".

structure which results from a choice of vocabulary, it is simply this: an arbitrary choice of a partial order on models.

However, one might consider the following response to these results: that we are not concerned with *all* possible vocabularies, but only with those that are, in some sense, natural; those that arise from our everyday use of language. If this response were valid, then it would have to be the case that the vocabulary given by natural language was quite restrictive: that we could not choose vocabularies arbitrarily, but that, for each situation, we naturally used a vocabulary which was tightly enough prescribed to fix our inferences adequately, and that we never, in the course of reasoning about a given situation, changed our vocabulary enough to change the validity of our inferences.

Davidson (following Austin (1956–7)) considers the following pattern of reasoning.

> 'I didn't know that it was loaded' belongs to one standard pattern of excuse. I do not deny that I pointed the gun and pulled the trigger, nor that I shot the victim. My ignorance explains how it happens that I pointed the gun and pulled the trigger intentionally, but did not shoot the victim intentionally. ... What is the relation between my pointing the gun and pulling the trigger, and my shooting the victim? The natural and, I think, correct answer is that the relation is that of identity. The logic of this sort of excuse includes, it seems, at least this much structure: I am accused of doing b, which is deplorable. I admit I did a, which is excusable. My excuse for doing b rests upon my claim that I did not know that $a = b$. (Davidson, 1980, p. 109)

What we are doing here is identifying two actions which are described differently, using different vocabulary: and the grounds on which we are identifying the actions are that they are the actions consist of the same sequence of events in the world. Thus, our everyday language exploits a wide range of redescriptions of the same action using different vocabulary. If all these things, variously redescribed, are the same action, then they must all have the same preconditions, the same effects, and they must all function in the same way in any adequate theory of action. And, finally, it is hard to see how the two actions of 'lifting the ball to position 2' and 'lifting the ball above positions 0 and 1, but not above position 2' could not be the same action. To establish identity, one needs, after all, only the facts of geometry, whereas for the identity between 'shooting' and 'pulling the trigger' one needs certain empirical matters, such as the presence of the bullet. So, if the identity holds in the case of Davidson's shooting example, it should, *a fortiori*, also hold in our example.

If, furthermore, we are to take seriously the appeal to natural language, we ought to take seriously Proposition 3. Natural languages typically abound in synonyms: they have large vocabularies, and most of the items in those vocabularies can be defined, in multiple ways, in terms of other items of those vocabularies. As the proposition shows, when we go about enlarging our vocabulary by definitions, we make our partial order successively more discrete. As the examples show, we do not, typically, need to make our vocabulary particularly large in order to make the resulting order completely discrete, and thus to make circumscription yield the same results as classical logic. And surely, if there is any particular vocabulary that the appeal to natural language ought to recommend, it must be the entire vocabulary of natural language, with its accompanying discrete partial order.

8 Conclusion

In these situation calculus inferences, the vocabulary plays a double role. One is its normal role: to form expressions in the relevant formal language. The other role, however, is to define the similarity relation between sets of fluents: two sets are more or less similar depending on inclusions between sets of fluents that have changed truth values in each of them. This relation is not independent of the vocabulary that it is formulated in; consequently, it is not actually a relation between possible worlds, but rather between *descriptions* of worlds. This means that the situation calculus, despite superficial similarities, is, in fact, markedly different from Lewis' semantics of counterfactual conditionals. It is also very different from the account of change that we get in physics: here we have a theory of physical change which relies on minimising a certain quantity (namely the integral of the action over a path in phase space), but the action is carefully defined so that the values of such integrals are independent of the coordinate system that they are expressed in.

One could argue, at this point, that these are little more than analogies: that the purpose of situation calculus is merely to imitate common-sense reasoning, and the role of change of vocabulary in such reasoning is small. Davidson's arguments show that this is not so: that change of vocabulary plays a very extensive role in our reasoning about actions.

These considerations have two main repercussions: theoretical and practical. Theoretically, it seems difficult to imagine how the situation calculus could be given a semantics in any strong sense. A semantics, however it might be expressed formally, should surely correlate linguistic items either with components of reality or with more abstract, mathematical objects: and, although the components of reality might have to be *described* in linguistic terms, the *correlation* itself should be independent of merely linguistic matters such as vocabulary. In particular, it should not be affected by the substitution of one vocabulary for another, logically equivalent one. And one should observe, in passing, that the rules of left logical equivalence and right weakening mean that any semantics for Kraus, Lehmann and Magidor-style inference relations must have these properties.

Practically, there are the following consequences. We would like to be able to use a theory (the situation calculus, maybe) by starting with whatever vocabulary presented itself and working with that. All we should have to worry about would be whether the vocabulary could describe the phenomena, and the correctness of our description: given both of these, the calculus ought to work. The situation ought to be similar to that in physics, where we can choose a coordinate system at our convenience: given a correct description, and correct equations, the results will be correct. But with the situation calculus as it is, we have more to worry about than merely adequacy and correctness: we also have to worry whether we have chosen a vocabulary which yields the right closeness relation between situations. All of our vocabularies are descriptively adequate, and our descriptions in any of them are correct. But they vary greatly in the predictions they make. And there seems to be no other way of telling whether the vocabulary is the magic correct one, other than by checking the results; and this rather diminishes the utility of the calculus as a predictive theory.

References

Austin, J.L. 1956–7. A Plea for Excuses. *Proceedings of the Aristotelean Society*, **57**. Reprinted in (Austin, 1970, pp. 175–204).

Austin, J.L. 1970. *Philosophical Papers*. Oxford University Press.

Costello, Tom, & McCarthy, John. 1999. Useful Counterfactuals. *Linköping Electronic Articles in Computer and Information Science*, **4**(12). Available at `http://www.ep.liu.se/ea/cis/1999/012`.

Davidson, Donald. 1980. The Logical Form of Action Sentences. *Pages 105–148 of: Essays on Actions and Events*. Oxford University Press. Originally published in N. Rescher (ed.), *The Logic of Decision and Action*, University of Pittsburgh Press, 1967.

Gabbay, Dov. 1985. Theoretical Foundations for Nonmonotonic Reasoning in Expert Systems. *In:* Apt, K. (ed), *Logics and Models of Concurrent Systems*. Berlin: Springer-Verlag.

Gabbay, Dov M., Hogger, C.J., & Robinson, J.A. 1994. *Handbook of Logic in Artificial Intelligence and Logic Programming*. Vol. 3. Clarendon. Volume Coordinator D. Nute.

Kraus, Sarit, Lehmann, Daniel, & Magidor, Menachem. 1990. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence*, **44**, 167–207.

Lewis, David. 1983a. Counterpart Theory and Quantified Modal Logic. *In:* (Lewis, 1983b). Originally appeared in *Journal of Philosophy* 65 (1968), 113–126.

Lewis, David. 1983b. *Philosophical Papers*. Vol. I. Oxford University Press.

Lewis, David. 1983c. Postscripts to 'Counterpart Theory and Quantified Modal Logic'. *In:* (Lewis, 1983b). Postscripts to (Lewis, 1983a).

Lewis, David. 1986. Counterfactuals and Comparative Possibility. *Pages 3–31 of: Philosophical Papers*, vol. II. Oxford University Press. Originally published in the *Journal of Philosophical Logic* 2 (1973): 418–46.

Lifschitz, Vladimir. 1994. Circumscription. *In:* (Gabbay *et al.*, 1994). Volume Coordinator D. Nute.

Makinson, David. 1989. General Theory of Cumulative Inference. *Pages 1–18 of:* Reinfranck, M. (ed), *Non-Monotonic Reasoning*. Lecture Notes in Artificial Intelligence, no. 346. Springer-Verlag.

Makinson, David. 1994. General Patterns in Nonmonotonic Reasoning. *In:* (Gabbay *et al.*, 1994). Volume Coordinator D. Nute.

Quine, Willard van Ormond. 1969. Speaking of Objects. *Pages 1–25 of: Ontological Relativity and Other Essays*. New York NY: Columbia.

Shanahan, Murray. 1997. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. MIT Press.