Ability and Action

Richmond H. Thomason
AI Laboratory
University of Michigan

October 4, 2002

Submitted to Sixth Symposium on Logical Formalizations of Commonsense Reasoning

Abstract

This is part of a larger project that is motivated in part by linguistic considerations and by the philosophical literature in action theory and the logic of ability, but that is also meant to suggest ways in which planning formalisms could be modified to provide an account of the role of ability in planning and practical reasoning.

In this version for the CSR workshop, I have suppressed most of the linguistic and philosophical issues, concentrating instead on the reasoning and its formalization.

1. Introduction

The literature on ability is scattered. I find it useful to divide it into the following five areas.

(1) **Philosophy of action.** Some of these philosophers (especially Kenny) were heavily influenced by Aristotle. Others (such as Richard Taylor) were working in a broader analytic tradition; most of these were concerned in one way or another with foundations of ethics, and especially with the free will problem.

(2) **Ordinary language philosophy.** The literature in the narrow ordinary language and Oxford traditions is centered around [Austin, 1956].

(3) **Analytic philosophy concerned with branching time, possible worlds and deontic logic.** This is not a unified tradition, although there are some coherent subgroups (like the group concerned with agency and the literature deriving from [Belnap, Jr. and Perloff, 1988]). For my purposes, by far the most important paper in this literature is [Cross, 1986], which in turn subscribes to the possible-worlds approach to conditionals, especially [Stalnaker, 1968].

(4) **The linguistics and philosophy of modal constructions.** [Kratzer, 1977] is an influential paper in this tradition.

(5) **Work in planning formalisms.** Ability is usually not addressed explicitly in the AI tradition, although intuitions about ability of course enter into the causal axioms for various actions. [Lin and Levesque, 1998] and [Meyer *et al.*, 1999] are exceptions.

2. Ambiguity Issues

Apparently, ‘can’ is ambiguous, as well as indexical or context-sensitive; the ambiguities may be manifold, and multiple dimensions of context may be involved. Any serious study of how to formalize ability has to begin by sorting out the ambiguities.

I believe that ‘can’ is ambiguous in several ways, and that only one sense is especially relevant for practical reasoning. As briefly as possible, I will try to indicate the readings of ‘can’ that aren’t relevant.

2.1. Possibility versus ability

(2.1a) My shoes can be under the bed.

(2.1b) It can be true that my shoes are under the bed.

(2.1c) ?My shoes are able to be under the bed.

(2.2a) I can prove that theorem.

(2.2b) ?It can be true that I will prove that theorem.

(2.2c) I’m able to prove that theorem.

The fact that Example (2.1b) is not unnatural, and is a paraphrase of Example (2.1a), and the fact that Example (2.1c) is unnatural, are indicators of a usage of 'can' to indicate that a possibility is not excluded. Contrast this with Example (2.2a–c), which is the usage I will be concerned with.

2.2. Generic versus occasional

(2.3) I can lift that rock.

(2.4) I can lift a 50 pound rock.

Example (2.4) is *generic*, it attributes a property to an agent that holds under a wide variety of times and circumstances—perhaps to all that are "normal" in some sense. I'm not concerned directly with generic uses of 'can', though I assume that to the extent that the meaning of a "generic tense" can be predicted from the meanings of the corresponding generic sentences, the following account may illuminate generic uses of 'can' as well. The occasional sense, which would be the most natural way to understand Example (2.3), is the one that I am concerned with here.

Generic abilities are not sufficient for planning purposes. I have a generic ability to climb trees, but of course I can't climb *any* tree. A three inch lodgepole pine with its lowest branches 30 feet above the ground is beyond my climbing abilities. Suppose that I'm out hiking and spot a grizzly a hundred yards away. Grizzlies are unpredictable, there is a potential emergency here, and I need a plan. I look around at the trees, I spot an alpine fir, and I say to myself "I can climb that tree." If the grizzly charges, this judgement has to ensure that I will get up the tree I selected. A 'can' that only guarantees that *I might* succeed in climbing it is not helpful here. I need to be sure that on this occasion, I *will* get up the tree if I try. A generic ability to climb trees is not what is wanted here. Even a generic ability to climb *this* tree is irrelevant, if the circumstances under which it can be expected to apply are not in place. For planning purposes, our judgments of ability have to provide successful results when put into practice.

In cases in which failures are noncritical, we can relax our standards somewhat; but even here, a judgment that a trial might fail tends to undermine the plan. Practical 'can's require success.

2.3. Further ambiguities

There is a further ambiguity or variability in the meaning of 'can' having to do with the status of the outcome. We often say that we can't do **a**, meaning not that a trial will fail to produce **a** but that it will produce consequences that are forbidden or undesirable; this is the sense in which I might tell someone I can't meet them at 2 because I have another appointment at that time. This is a natural extension of 'can' for planning purposes, and I believe that it fits well into the theory that I will develop. But I will not have more to say about it, at least in this draft.

3. Possible worlds semantics for 'can'

[Cross, 1985] is a good beginning in investigating the semantics of 'can'. Like possible worlds theories of the conditional, Cross' account exploits a function that, given a clause ϕ and an agent A, selects a set of possible worlds that is in some sense "close" to the actual world w. Intuitively, these are the worlds that provide appropriate test conditions for A's "performance" of ϕ, or rather (since ϕ expresses a proposition, rather than an action) for A's taking steps to bring about an outcome in which ϕ holds. Cross' semantic rule for $<A>$ is this.

(Can1) $M \models_w <A>\phi$ if and only if $M \models_{w'} \phi$ for some $w' \in g(\phi, a, w)$.

$<A>\phi$ is true at a world w if and only if ϕ is true at some world w' in $g(\phi, \mathbf{A}, w)$. The function g selects the set of worlds that, relevant to the circumstances in w, would provide appropriate "test conditions" for, as Cross puts it,

(3.1) testing whether the truth of ϕ is within $\mathbf{A}$'s abilities in w.

The function meets the following two conditions.

(SB) If $\{w\colon M \models_w \phi\} \subseteq \{w\colon M \models_w \psi\}$ then $g(\phi, \mathbf{A}, w) \subseteq g(\psi, \mathbf{A}, w)$.
(AC) If $M \models_w \phi$ then $w \in g(\phi, \mathbf{A}, i)$.

According to Cross, $<A>\phi$ is true at w if and only if for some $w' \in g(\phi, \mathbf{A}, w)$, ϕ is true in w', where $<A>\phi$ is proposed as an adequate formalization of a sentence involving the application of 'can' to a subject formalized by A and a clause formalized by ϕ. This makes 'A can' a relativized modal possibility operator that is closely related to the relational operator $\Diamond\!\!\rightarrow$ of [Lewis, 1973], which David Lewis proposed as a formalization of conditional 'might' constructions.

The relativization solves some of the obvious problems of using a standard, nonrelativized possibility operator to formalize 'A can', such as Kenny's objection ([Kenny, 1976b, Kenny, 1976a] that $\boldsymbol{Can}_A[\phi \vee \psi]$ does not entail $\boldsymbol{Can}_A\phi \vee \boldsymbol{Can}_A\psi$.[1] Also, the idea that the meaning of 'A can' is associated with a hypothetical test in which $\mathbf{A}$ is given a fair chance, under normal circumstances, at an attempt to perform an appropriate action, is very appealing. However, Cross' proposal is unintuitive in some important respects.

4. The clausal argument of 'can'

Cross is working within a logical framework in which actions are not available. In many other cases (deontic logic, for example) the conflation of propositions and actions, even if it is unintuitive, does not prevent the development of sophisticated formalisms that illuminate the logical issues. The same may be true here. However, in a formalism that does provide for action, it would be more natural to construe $\boldsymbol{Can}_A$ as an operator on actions.

The sentential formalism may well involve serious conceptual errors. But instead of trying to expose these, I will try to show that making actions the arguments of $\boldsymbol{Can}_A$ provides a more detailed and robust account of the selection function g.

[1] For the moment, I'll use $\boldsymbol{Can}_A$ for an unformalized representation of the 'can' of ability.

5. Is ‘can’ a possibility operator?

To put it roughly, Cross’ theory of the ‘can’ of ability is based on an equivalence between ‘I can’ and ‘If I tried I might’. I reject this equivalience: it seems to me that ‘If I tried I would’ is a more intuitive conditional explication.

At this point, I will skip some detailed discussion of this issue. See the version of the paper at www.umich.edu˜rthomaso/documents/action/ability.ps for details.

To sum the matter up, the evidence concerning the modal status of ‘can’ appears to be mixed, making it difficult to provide a theory that is unequivocally supported by the evidence. However, I believe that the following theory is well enough supported to be plausible, and that the apparent counterexamples can be explained away in a principled way.

6. A conditional theory of ability

The logical situation with respect to ability is, I think, somewhat similar to the one that prevails with conditionals. According to the *variably strict* theories ([Lewis, 1973] is an example), a conditional $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \psi$ is true in case ψ is true in every one of a set of worlds depending on ϕ. According to the *variably material* theories ([Stalnaker, 1968, Stalnaker and Thomason, 1970] are examples) $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \psi$ is true in case ψ is true in *a single* world depending on ϕ. The chief difference between the two is that conditional excluded middle

(6.1) $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ [\psi \vee \chi] \rightarrow [\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \psi \vee \boldsymbol{If}\ \phi\ \boldsymbol{then}\ \chi]$

holds in the variably material accounts.

The variably strict theories are much more popular; the variably material theories seem to be much better supported by the linguistic evidence.

Cross’ rule represents a variably strict theory of $\boldsymbol{Can}_A$ (or rather, of its negation, $\neg\boldsymbol{Can}_A$). The corresponding variably material theory would make $g(\phi, \mathbf{A}, w)$ a unit set. We can simplify the picture by positing a function f from formulas, agents and worlds to worlds. The satisfaction condition for ability would then be as follows.

(6.2) $M \models_w \boldsymbol{Can}_A\phi$ if and only if $M \models_{f(\phi,\mathbf{A},w)} \phi$.

The intuitive meaning of f is that $f(\phi, \mathbf{A}, w)$ should be the closest world to w in which $\mathbf{A}$ tries to bring about ϕ.[2] We impose one condition on f.

(6.3) If $M \models_w \phi$ then $f(\phi, A, w) = w$.

Rule (6.2) has the effect of validating $\boldsymbol{Can}_A\phi \vee \boldsymbol{Can}_A\neg\phi$; Condition (6.3) validates $\phi \rightarrow \boldsymbol{Can}_A\phi$. The logic of $\boldsymbol{Can}_A$ can be axiomatized by adding these as axiom schemas to basic axioms for modal operators. One could ask whether the underlying modal logic should be that of **S4** or even **S5**, but I think it is unrewarding to press these details too far. Instead, I will simply note that the reformulated modal theory of ability is closer to situation-based accounts, because f is now analogous to the Result function.

[2] Cross requires that his function g should give the attempt a fair trial—things must be normal with respect to A’s attempt to bring ϕ about. I agree that this condition is important for generic ability, but that it is not part of occasional ability. Practical ability has to take adverse circumstances into account.

7. Providing ‘can’ with actions as arguments

This project divides into two parts: (i) adopting a logic with a quantificational domain that contains actions (here, I mean individual actions, not action types) and (ii) providing an account of the formalization of commonsense examples involving ‘can’.

Two quite different traditions do the first job: eventuality-oriented natural language semantic theories, such as [Parsons, 1990, Higginbotham *et al.*, 2000], and action centered dynamic theories such as the Situation Calculus. [Steedman, 1998, Steedman, 1998] combine both traditions. I can rely on the natural language tradition to deal with the connections to natural language issues, and will concern myself from now on with task (i). I will try to focus on reasoning issues in the remainder of this paper, and will mainly draw on the dynamic tradition.

The account that is formulated in the following sections will be simplified in various ways. Eventually, I would like to produce a formulation that takes into account (1) full abstraction hierarchies of higher and lower-level actions, that is, hierarchies of actions and hierarchical planning; (2) aspectual types and actions that are characterized in terms of the results they achieve; and (3) concurrence.

I will use the Situation Calculus as a starting point. This too could be considered to be a simplification; but the Situation Calculus has proved to be remarkably robust. Issues about ability are closely related to knowledge-how, but in this version I will not say much about this aspect of things, which is complicated in its own right. The idea of incorporating knowledge-how into a planning formalism goes back to [Moore, 1985].

8. Situation Calculus

I would like to approach the problem of formalizing ability in a way that could be carried out in most formalisms for reasoning about action and change. But it will be convenient to explain the basic ideas using a version of the Situation Calculus.

I’ll use Many-Sorted First-Order Logic as the vehicle of formalization. There is a sort AC of actions, a sort FL of fluents or states, a sort SI of situations, and a sort IN of garden-variety individuals. I’ll adopt a convention of flagging the first occurrence of a sorted variable in a formula with its sort. I’ll do this with constants also, except that: (1) a is reserved for constants of sort AC, (2) s is reserved for constants of sort SI, (3) f is reserved for constants of sort FL, and (4) c is reserved for constants of type IN.

In the standard SC approach, you have a function $\mathbf{r}$ from actions and situations to situations. $\mathbf{r}(\mathbf{a}, \mathbf{s}) = \mathbf{s}'$ means that $\mathbf{s}'$ is the situation that results from performing $\mathbf{a}$ in $\mathbf{s}$. (The existence of $\mathbf{r}$ presupposes a deterministic sort of change, at least as far as action-induced change goes.) Our formal language contains a function letter RESULT denoting the function $\mathbf{r}$.

If you suppress considerations having to do with causality and the Frame Problem (which I propose to do for the time being) the formalism is pretty simple. Planning knowledge is indexed to actions, in the form of *causal axioms* which associate conventional effects and preconditions with actions. The causal axiom for an action $\mathbf{a}$ denoted by a has the following form.

(8.1) $\forall x_s[\textsc{Pre}(a,x) \rightarrow \textsc{Post}(\textsc{Result}(a,x))]$.[3]

Here, $\textsc{Pre}$ is the precondition for a and $\textsc{Post}$ is the postcondition or effect of a.[4]

9. Situation Calculus with Explicit Ability

Often, a causal axiom of the form Condition (8.1) is read: "If a is done and $\textsc{Pre}(a,s)$ is true, then $\textsc{Post}(a,\textsc{Result}(a,s))$ is true." This provides a perfectly satisfactory basis for reasoning with actions and plans as long as one is only interested in the successful performance of actions. But it is counterintuitive when it may be important to reason about unsuccessful "performances"—i.e., about attempts to perform an action which may fail. This is exactly the sort of reasoning in which "trying" is invoked in informal, commonsense reasoning.

Consider, for instance, a case in which I want to talk to my wife on the telephone. I have a standard method of trying to talk to anyone on the telephone, which consists in (1) finding out the telephone number in case I don't know it, and proceding to step (2) otherwise; then (2) locating a telephone in case one isn't handy, and proceding to step (3) otherwise; and then (3) dialing the phone number. This method will succeed if my telephone is working and my wife is not using her telephone and is able to answer it; that is, I can call her if these conditions are satisfied. I may know that I can talk to her and proceed to try to talk to her armed with this knowledge; but we seldom act with this sort of assurance. Much more typically, I may not know that I can talk to her but will try to talk to her simply assuming that I can do so, and without any clear fallback plan in case of failure.[5]

To formalize these ideas in a Situation Calculus format, we have to get explicit about trying. This means that we need a more elaborate theory of actions.

9.1. Trying

Many common-sense actions[6] are described in terms of the states they achieve; linguists use the term 'telic' for such actions. Generalizing the idea that emerged from the telephoning example, I want to say that the ability to perform such actions consists in (1) having a salient method of achieving the goal of the action in one's repertoire, and (2) being in circumstances such that if the method were acted on, the goal would be achieved.

Although for nontelic actions (jumping, for instance) it is not clear that a method is involved, we do need to invoke a method in some cases; this means that we will have to incorporate some of the ideas of hierarchical planning in the formalism. To simplify things, I'll assume that the hierarchy has a depth bounded by 2. The models of SC that I have in

[3] To simplify notation, I am write $\textsc{Pre}(a,s)$ rather than $\textsc{Holds}(\textsc{Pre}(a),s)$. Uses of $\textsc{Holds}$ and $\textsc{Value}$ are suppressed in similar ways throughout the rest of this paper.

[4] In planning formalisms, what I am calling preconditions are often separated into two kinds of conditions: the ones that are under the agent's control (which are often called "preconditions") and the ones that are not (which are often called "constraints". I do not distinguish between these two sorts of conditions here.

[5] Typically, people will act on commonsense plans without anything like a proof that they will succeed. All that is seems to be required is the absense of a reason for their failure and relatively low risk or cost of trying to achieve the goal, in relation to the benefits of achieving it.

[6] For instance telephoning someone, crossing a street, opening a door, turning on a light.

mind,then, will involve a function **Method**. This function inputs an action **a** and a situation **s**, and outputs the agent's "method of choice" for **a** in **s**. This method is a composite action whose components are basic actions; **a** is basic in case for all **s**, $\mathbf{Method}(\mathbf{s}, \mathbf{a}) = \mathbf{a}$.[7]

I am making a number of assumptions here; let me explain them briefly. (1) The constraint on the values of **Method** formalizes the idea that the depth of the abstraction hierarchy on actions is at most 2. (2) An agent may have many ways of performing **a** in **s**. I am assuming that one of these is the one that the agent *would* use if it tried to perform **a**. So **Method** has a certain amount of counterfactual content. (3) An agent may have no way of performing **a** in **s**. In the present draft, I'll deal with this in the simplest possible way, by (i) introducing a null act, NULL, with the understanding that $\mathbf{Method}(\mathbf{a}, \mathbf{s}) = \textsc{Null}$ means that **Method** is undefined for **a** in **s**. (4) We want to avoid having to apply the **Method** function to composite actions; so we assume that **Method** is a homomorphism with respect to composition: $\mathbf{Method}(\mathbf{a};\mathbf{b}, \mathbf{s}) = \mathbf{Method}(\mathbf{a}, \mathbf{s});\mathbf{Method}(\mathbf{b}, \mathbf{s})$.

Whatever "having a method of choice for performing **a** in **s**" means, it is close to saying that an agent knows how to **a** in **s**; knowing how and ability are intimately connected. I won't explore that connection here.

The **Method** function seems to incorporate an insight from hierarchical planning. But the people who formalize hierarchical planning domains do not distinguish between ways of *trying* to do something and alternative methods of doing something, lumping the two together. I can call my wife by calling her on a cell phone or a pay phone; but these are ways of calling her, not of trying to call her. Roughly, a method of trying to do something has to be a way of doing it that is appropriately specific.

We denote the function **Method** by a function letter METHOD of the planning formalism. With the incorporation into the formalism of a distinction between an action **a** and the action of trying to do **a** in **s**, we can revise the causal axiom of the classical Situation Calculus for a constant a denoting an action as follows: we are interested in the results of *trying* to do an action, rather than the results of doing the action itself.

(9.1) $\forall x_s[\textsc{Can}(a, x) \rightarrow \textsc{Post}(\textsc{Result}(\textsc{Method}(a, s), x))]$.

Such axioms correspond to platitudes of the sort

(9.2) If I can open the door, then after I try to open the door the door will be open.

$\textsc{Can}(a, s)$ can now be characterized in terms of preconditions and constraints on the action denoted by a. For instance, suppose that a denotes a basic action **a**, and that s denotes **s**. Then the success conditions for **a** in s, i.e. the conditions under which trying to perform **a** (in this case simply performing **a**) will achieve the effects conventionally associated with a are simply the preconditions of **a**.

(9.3) $\textsc{Can}(a, s)$ amounts to $\textsc{Pre}(a, s)$.

[7] I use the term 'basic' because actions that are fixpoint of **Method** are similar to the so-called "basic actions" that have been discussed in the literature on the philosophy of action. See, for instance, [Brand, 1984, Care and Landesman, 1968, Danto, 1973, Goldman, 1970].

For basic actions, then, we are merely repackaging the usual format for causal axioms.

But things are more interesting if a denotes a nonbasic action. For example, a_0 could denote opening a certain door, and $\textsc{Method}(a_0, s)$ could denote grasping the doorknob, exerting a clockwise turning force on the doorknob, and pushing against the doorknob. Let a_1, a_2 and a_3 denote these three actions. I want the preconditions of pushing against the doorknob in this case to include the door's not being stuck. But these are not preconditions of pushing against the doorknob considered in itself; they only count as preconditions when it is considered as a part of $\mathbf{a}_0$. To deal with this complication, I will generalize $\textsc{Pre}$ to involve three arguments: $\textsc{Pre}(a, a', s)$ combines the preconditions of $\mathbf{a}$ in $\mathbf{s}$ when it is considered as a component of $\mathbf{a}'$. Suppose that $\textsc{Pre}(a_1, a_0, s)$ is 'The agent is next to the door', that $\textsc{Pre}(a_2, a_0, s)$ is 'The door isn't locked', and that $\textsc{Pre}(a_3, a_0, s)$ is 'The door isn't stuck'.[8] Then, in this case, $\textsc{Can}(a_0, s)$ amounts to 'The door is neither stuck nor locked'.

In this simplified treatment there is a straightforward definition of $\textsc{Can}$:

(9.4) $\textsc{Can}(a_0^s, s)$ amounts to $\bigwedge\{\textsc{Pre}(a_i, a_0, s) :$ a_i is a component of $\textsc{Method}(a_0, s)$.

The most interesting new arenas for formalization that arise in SC with ability have to do with axiomatizing domain knowledge about what can be done. (Formalizing an abstraction hierarchy for actions and the **Method** function in SC is not trivial, but since fairly complex domains have been formalized in connection with hierarchical planning, we can assume, I think, that this part of the formalization is feasible.) Often, we don't know that we can do things, but it is reasonable to assume that we can—assuming that a door I am about to open is a typical case of this kind. In formalizing ability, then, we can expect to use defaults liberally.

9.2. An example sketch

There isn't space here to work up an example with full explicitness. Instead, I will take a case that isn't entirely trivial and explain enough about the formalization to make it plausible that we can go from the common sense knowledge we typically have to axioms.

Assume that I have come up with the following method for achieving a goal:

1. Go to the bookcase.
2. Find my copy of *Emma*.
3. Take it off the shelf.
4. Go to the chair.
5. Sit down.
6. Turn on the lamp switch.
7. Read chapter 1 of the book.

The following paragraphs are meant to show that the appropriate knowledge is readily available, and in many cases is available in fairly general form.

[8] I'm imagining a door with only one lock, which works by preventing the doorknob from turning.

(1) I am assuming that this scenario takes place in my house. I know the layout of my house. Part of knowing the layout is knowing whether paths are open; assuming it's a normal house, this is just a matter of knowing which passages are closed. Having a method of going from my current location to the bookcase is a matter of finding a preferred route. This too can be done from a spatial representation of the house. If my preferred route route from my location to the bookcase is open, it follows (by default) that I can go to the bookcase.

(2) If I have a recognition criterion for an object and it is in a reasonably small area (like a bookcase) and I am at that area, it follows (by default) that I can find it.

(3) If I have found a book-sized object on an open shelf and and I'm next to it then (by default) I can pick it up.

(4) This is like (1).

(5) If I am at a chair and it is not occupied, then (by default) I can sit in it. Either I remember that this chair is unoccupied (by people or things), or I assume (by default) that it is unoccupied.

(6) This case is typical about ability-related assumptions about the workings of artifacts. In the context of this plan, I am assuming that the preconditions for turning on the lamp are inherited by turning on the switch. If the switch is working, the device is plugged in, the bulb is not broken (I assume a lamp with one bulb), the bulb is screwed in completely, and the power is on, the lamp will go on if I turn the switch. I assume the first three of these by default, and either I assume the fourth as well, or (more usually) I know it from sensory information.

(7) If we don't take into account resource constraints having to do with available time, this is a matter of standing know-how (I know how to read English, Emma is written in English), physical ability (I can read 10 point type if I'm wearing my glasses, this copy of *Emma* is in 10 point type), and—another artifact assumption—all the pages of chapter 1 are in this copy of the book. (We all have experienced defectively made copies of books, so this last assumption is a default,)

It is easy to see that the appropriate defaults are formalizable. The following circumscription-style axiomatization of the ability requirements of the lamp illustrate the pattern.

$$\forall x \forall y_s [[\mathit{Lamp}(x) \wedge \neg \mathit{Ab}_1(x,y)] \rightarrow \mathit{Working}(\mathit{switch}(x,y),y)]$$

$$\forall x \forall y_s [[\mathit{Lamp}(x) \wedge \neg \mathit{Ab}_2(x,y)] \rightarrow \mathit{Plugged\text{-}in}(x,y)]$$

$$\forall x \forall y_s [[\mathit{Lamp}(x) \wedge \neg \mathit{Ab}_3(x,y)] \rightarrow \neg \mathit{Broken}(\mathit{bulb}(x,y),y)]$$

$$\forall x \forall y_s [[\mathit{Lamp}(x) \wedge \neg \mathit{Ab}_4(x,y)] \rightarrow \mathit{Screwed\text{-}in}(\mathit{bulb}(x,y),y)]$$

This axiomatization style provides a home for a more elaborate theory of abnormalities. Lamps for sale in a store may not be plugged in, so they suffer from the second form of anomaly. If my house is regularly cleaned by someone who unplugs the lamp, then the lamp in the example is regularly anomalous. Expectations about normality may be suspended in some cases for reasons that have more to do with risk reduction than epistemology: this is what happens during a pre-flight check.

9.3. Advantages and uses

An abnormality theory axiomatizing the circumstances under which an executed plan may fail can serve to diagnose and explain these failures. To the extent that the abnormality theory is complete, we can expect an abnormality to hold when we experience such frustrations. In special cases, we may be able to design abductive algorithms that produce the set of anomalies that could explain a failure. (See, for instance, [Kautz, 1990], [Eiter, 2002], [Lin and You, 2002].) With additional knowledge, we may even be able to rank the anomalies in order of plausibility.

Such a theory provides the beginnings, anyway, of a formal account of the knowledge that is used in producing excuses. As J.L. Austin showed in [Austin, 1956–57], the language of excuses is extraordinarily rich and subtle—this is a strong indication that it corresponds to a well-developed area of commonsense reasoning.

Ethical applications aside, the ability to reason about courses of action that have failed is crucially important for practical reasons. The modification of planning formalisms that is suggested here seems to provide a place for this sort of reasoning in a way that matches to some extent the common-sense language and organization of the relevant knowledge.

I think it is an interesting and important feature of this project that plan abstraction hierarchies occupy a central place in the theory. It is impossible, I believe, to begin to give an adequate account of execution failure without enriching the theory of action with a theory of trying, and this leads at once to a hierarchical formulation. Hierarchical representations are, of course, often used in planning applications, but I believe that they are usually motivated by reasons having to do with modularity and efficiency, of the sort that are usually advanced in connection with abstraction hierarchies. This suggests that plan hierarchies are in principle, at least, dispensible in formalizing reasoning about action, and in fact the ideas of hierarchical planning are not usually incorporated in logical formalizations of planning.

I want to suggest that the more formal work has been able to avoid a hierarchical formulation only because it has used an oversimplified model of action that doesn't apply well to reasoning about the failure of actions. The fact that plan hierarchies provide the missing element is very welcome, since hierarchical planning is widely used, and many complex domains have been formalized. Even if the formalizations are not fully declarative and logical, this work shows that the relevant knowledge is there and can be represented.

Although the difference between the two formalisms makes a direct comparison difficult, the account of 'can' that emerged from this simple treatment in the Situation Calculus is similar to the more abstract, conditional account of Section 6, In both cases, to say that an agent can perform an action provides a condition that ensures the successful performance of the action. The advantage of the SC treatment is that it provides a represention from which we can recover explicit conditions of success. Nothing, of course, guarantees that these conditions should be anything that the agent can control or even know—in formalizing actions that depend on an element of luck, we may have to resort to unknowable "hidden variables." But in the cases where classical planning algorithms are appropriate, it seems that we can recover useful conditions.

10. Conclusion

In this draft, I have only tried to formalize the simplest version of an action-and-change formalism capable of explicitly representing ability. These simplifications need to be replaced with something more adequate.

Also, the work of integrating the sort of approach advocated here into logical formalisms for reasoning about action and time remains to be done. I haven't begun to work out things like the relation of ability to things like knowing how, nonmonotonicity in action and change, causality, and eventualities of different aspectual types.

However, I hope that this presentation makes a plausible case for the value of the modifications that are suggested here in accounting for commonsense reasoning about plan execution and plan failure.

References

[Austin, 1956] John L. Austin. Ifs and cans. *Proceedings of the British Academy*, pages 109–132, 1956.

[Austin, 1956–57] John L. Austin. A plea for excuses. *Proceedings of the Aristotelian Society*, 57:1–30, 1956–57.

[Belnap, Jr. and Perloff, 1988] Nuel D. Belnap, Jr. and Michael Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54:175–199, 1988.

[Brand, 1984] Myles Brand. *Intending and Acting: Toward a Naturalized Action Theory.* The MIT Press, Cambridge, Massachusetts, 1984.

[Care and Landesman, 1968] Norman S. Care and Charles Landesman, editors. *Readings in the Theory of Action.* Indiana University Press, Bloomington, Indiana, 1968.

[Cross, 1985] Charles B. Cross. *Studies in the Semantics of Modality.* Ph.d. dissertation, Philosophy Department, University of Pittsburgh, Pittsburgh, Pennsylvania, 1985.

[Cross, 1986] Charles B. Cross. 'Can' and the logic of ability. *Philosophical Studies*, 50:53–64, 1986.

[Danto, 1973] Arthur Coleman Danto. *Analytical Philosophy of Action.* Cambridge University Press, Cambridge, 1973.

[Eiter, 2002] Thomas Eiter. On computing all abductive explanations. In Rina Dechter, Michael Kearns, and Richard S. Sutton, editors, *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 62–67, Menlo Park, California, 2002. American Association for Artificial Intelligence, AAAI Press.

[Goldman, 1970] Alvin I. Goldman. *A Theory of Human Action.* Princeton University Press, Princeton, New Jersey, 1970.

[Higginbotham *et al.*, 2000] James Higginbotham, Fabio Pianesi, and Achille C. Varzi, editors. *Speaking of Events.* Oxford University Press, Oxford, 2000.

[Kautz, 1990] Henry A. Kautz. A circumscriptive theory of plan recognition. In Philip R. Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 105–133. MIT Press, Cambridge, Massachusetts, 1990.

[Kenny, 1976a] Anthony Kenny. Human ability and dynamic modalities. In Juha Manninen and Raimo Tuomela, editors, *Essays on Explanation and Understanding*, pages 209–232. D. Reidel Publishing Company, Dordrecht, 1976.

[Kenny, 1976b] Anthony Kenny. *Will, Freedom, and Power.* Barnes and Noble, New York, 1976.

[Kratzer, 1977] Angelika Kratzer. What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1(3):337–356, 1977.

[Lewis, 1973] David K. Lewis. *Counterfactuals.* Harvard University Press, Cambridge, Massachusetts, 1973.

[Lin and Levesque, 1998] Fangzhen Lin and Hector J. Levesque. What robots can do: Robot programs and effective achievability. *Artificial Intelligence*, 101(1–2):201–226, 1998.

[Lin and You, 2002] Fangzhen Lin and Jia-Huai You. Abduction in logic programming: A new definition and an abductive procedure based on rewriting. *Artificial Intelligence*, 140(1–2):175–205, 2002.

[Meyer *et al.*, 1999] John-Jules Ch. Meyer, Wiebe van der Hoek, and Bernd van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113(1–2):1–40, 1999.

[Moore, 1985] Robert C. Moore. A formal theory of knowledge and action. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex Publishing Corporation, Norwood, New Jersey, 1985.

[Parsons, 1990] Terence Parsons. *Events in the Semantics of English: a Study in Subatomic Semantics.* The MIT Press, Cambridge, Massachusetts, 1990.

[Stalnaker and Thomason, 1970] Robert C. Stalnaker and Richmond H. Thomason. A semantic analysis of conditional logic. *Theoria*, 36:23–42, 1970.

[Stalnaker, 1968] Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell Publishers, Oxford, 1968.

[Steedman, 1998] Mark Steedman. The productions of time. Unpublished manuscript, University of Edinburgh. Available from http://www.cogsci.ed.ac.uk/~steedman/papers.html., 1998.