Ability, Action, and Context

Richmond H. Thomason
Philosophy Department
University of Michigan

October 31, 2005

Note. This is a working draft. The ideas presented here may not be in their final form.

Abstract

This paper proposes a formalization of ability 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 match well with planning formalisms, and so to provide an account of the role of ability in practical reasoning. Some of the philosophical literature concerning ability, and in particular [Austin, 1956], suggests that some ways of talking about ability are context-dependent. I propose a way of formalizing this dependency.

1. Introduction

The philosophical and linguistic literature on ability is by now fairly considerable, but it divides into several disparate projects and themes, and there has not been much communication between the subcommunities that have pursued these themes. This paper represents the first stage of a project that will try to canvass this material, extracting the insights that seem to be most valuable and permanent, and to develop a theory that relates ability to actual reasoning problems by integrating it in a planning formalism of the sort deriving from [McCarthy and Hayes, 1969].

This report is part of a projected larger work, in which I will be drawing on a literature that spans the last fifty years or so. I classify the authors who produced this work into the following five groups.

(1) **Philosophers 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 philosophers.** The literature in the narrow ordinary language and Oxford traditions is centered around [Austin, 1956].

(3) **Philosophers interested in 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) **Linguists and philosophers interested in modal constructions.** [Kratzer, 1977] is an influential paper in this tradition.

(5) **Computer scientists interested in foundations of planning.** As far as I know, this group has not had much to say about ability. Their theories tend to scatter ability around in the antecedents of causal axioms for various actions, But [Lin and Levesque, 1998] and [Meyer *et al.*, 1999] are recent exceptions.

Here I will concentrate mainly on logical foundations, so the most relevant of the areas listed above are (3) and (5). But even for the more limited project, all of this work has some relevance, and from time to time I will mention papers from (1), (2), and (4).

[Cross, 1986], based in part on [Cross, 1985], contains a number of important insights into the logical problem of formalizing ability, and I take this work as my starting point.

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 its meaning has to begin by taking up a position on these issues.

I rely on linguistic tests for ambiguity, and base hypotheses about sensitivity to context on somewhat less objective considerations; if the variation in meaning appears to be continuous and the theory of context can be stated precisely and is more or less plausible, I'm willing to entertain a contextual theory.

Deploying ambiguity tests from [Thomason and Stalnaker, 1973], [Cross, 1986, pp. 54–57] provides convincing evidence for at least one ambiguity in 'can'. According to these tests, one sense of 'can' should be formalized as a sentence modifier and the other should be formalized as a predicate modifier. The simplest of the tests involves the following paraphrase.

(2.1a) Linguistic traits can diffuse across an area.

(2.1b) It can be true that linguistic traits will diffuse across an area.

(2.2a) I can prove that theorem.

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

The fact that Example (2.1b) is not unnatural, and is a paraphrase of Example (2.1a), suggests that 'can' is used as a sentence modifier in Example (2.1a). The unnaturalness of Example (2.2b), and the fact that if it makes sense at all it forces a reading of Example (2.2a) according which it merely expresses the possibility of an outcome in which I prove the theorem, suggests that Example (2.2a) is more naturally interpreted as a predicate modifier.

Cross correlates the ambiguity supported by this test with a sense of ‘can’ that attributes an “ability” to an agent. I agree, but I need to explain some details and make some hedges.

Enabling factors. I want to construe ‘ability’ very broadly as far as enabling factors are concerned. In general, ability can depend on favorable circumstances, on the presence of appropriate knowledge, and on nonepistemic properties of the agent. I can truly say

(2.3) I can’t write a check.

either because my bank balance is negative, or because I don’t know where my checkbook is, or because my hand is injured. I believe that the same sense of ‘can’ is involved in each case.

Occasional and generic readings. An example like

(2.4) I can lift that rock.

attributes a time-bound, circumstance-bound state to an agent. As usual with such attributions, there are corresponding, related generic sentences. An example like

(2.5) I can lift a 50 pound rock.

would be most plausibly understood as 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.[1] 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.

Generic uses of simple sentences without modal auxiliaries are marked by present tense, as in the following pair of examples.

(2.6a) I am walking to work.

(2.6b) I walk to work.

Example (2.6a) is not generic—it refers to a particular occurrence of walking to work. In contrast, Example (2.6b) deals with a regularity rather that an individual event; it says that I normally or generally walk to work.

I assume that

(2.7a) [I borrowed an alarm clock, so tomorrow] I can get up at 5.

(2.7b) [I wake up really early, so in particular] I can get up at 5.

[1] See [Carlson and Pelletier, 1995] for discussion of generics.

stand in a relation that is similar to the one between Example (2.6a) and Example (2.6b). In this case, the presence of a modal auxiliary ('can') precludes the use of the distinction between present progressive and simple present to mark the difference between occasional and generic uses. In the absense of syntactic differentiation, I have used context (in square brackets) to produce the two uses.

Example (2.7a) refers to a particular occurrence of a state of being able to get up at 5, just as Example (2.6a) refers to a particular occurrence of walking to work. Example (2.7b) refers to a general, standing state of being able to get up at 5, just as Example (2.7b) refers to a general, standing preference for walking as a method of transportation to work.

Since generic abilities can refer to highly qualified and specific states—I could, for instance, talk about my generic ability to hike the Rumble Creek trail when I've gotten in shape and the snow has melted and the weather is good and the brush isn't too thick—attributions of generic abilities can blend into ones of occasional abilities and the distinction between the two can be hard to make in practice.

Nevertheless, I believe that the difference is crucial in understanding the logic of ability. We have a very robust intuition that abilities are important in planning; in most practical cases when we are concerned to know whether we can do something, it is so that we can fit it into a plan.

Generic abilities are not sufficient for such 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.

The position I have reached can be summarized in two parts.

(i) There are occasional and generic uses of 'can'. I assume that the semantic relation between these uses is similar to the relation between other occasional and generic uses, and in particular that the occasional uses are fundamental, whereas the generic ones are derived by whatever general process relates generic uses to occasional ones.

(ii) The fact that generic uses of 'can' are not syntactically differentiated from occasional uses is an obstacle in developing a semantic theory of ability. Generic sentences are difficult to falsify. In extreme cases, a sentence like

(2.8) I eat meat.

can boil down to just

(2.9) There are some occasions on which I eat meat.

The same applies to sentences attributing generic ability, like the following example.

(2.10) I can distinguish Coca-Cola from Pepsi-Cola.

The definitive test of a generic ability claim is like that of an occasional ability claim and a conditional—you give it a fair trial and see what happens. In the case of Example (2.10) (unless I'm joking) I mean that I can differentiate by taste, so the trial would consist in presenting me with blindfold examples. If I get it wrong, I can always give an excuse, and ask for another test. If I get it right on the next ten trials, you might be inclined to agree with me that Example (2.10) is true; but it isn't at all clear what level of performance over a number of trials would make Example (2.10) false.

This fact, together with the absense of any linguistic difference between the generic and occasional uses of 'can', together with the existence of a sense of 'can' ('can be true') that is in fact a modal possibility operator, can make a similar theory of the 'can' of ability attractive. In fact, all the formal theories of 'can' that I know of treat it as a species of possibility operator. I want to reject such an account, and I hope that I have persuaded anyone who has read this far to at least suspend the assumption that 'can' is some sort of possibility operator.

In nondeterministic settings, we can't in general be sure that trying to do something will bring it about. I can try to dry my laundry by putting it out to dry, but this will only work if the weather cooperates. If I'm right about the logic of 'can', we often do things without having been sure in advance that we could do them; when these things were planned or intended, we could not have known that the plans were feasible. I'm prepared to accept this conclusion; when we are faced with uncertainty, we often have to act under uncertainty. We form plans based on more or less well-founded assumptions or even on hopes—and of course practical commitment to such plans can be risky. I believe that practical reasoning under uncertainty is compatible with the account of 'can' I am advocating here, as long as the truth of statements involving the practical 'can' depends, among other things, on the way the future unfolds.[2] However, I will not have much to say (in this draft at least) about the nondeterministic case.

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 α, meaning not that a trial will fail to produce α 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.[3] I will return to this issue briefly in Section 11, below.

3. Possible worlds semantics for 'can'

Like possible worlds theories of the conditional, Cross' theory 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

[2] I am thinking of a semantics for 'can' like the semantics for future tense in [Thomason, 1970].

[3] Desirability, as well as feasibility, can serve as a filter in plan generation and plan maintenance. See, for instance, [Pollack and Horty, 1999].

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.

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

$<a>\phi$ is true at a world w if and only if ϕ is true at some world w' in $g(\phi, a, w)$. 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 a's abilities in w.

The function meets the following two conditions.

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

According to Cross, $<a>\phi$ is true at w if and only if for some $w' \in g(\phi, a, w)$, 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.

According to this theory, 'a can' is 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 $\boldsymbol{Can}_a\phi \vee \boldsymbol{Can}_a\psi$.[4] Also, the idea that the meaning of 'a can' is associated with a hypothetical test in which 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 respects, which I will examine separately in the following sections.

4. The clausal argument of 'can'

Cross relies on the 'It is X-ly true that ϕ' test, illustrated in Examples (2.1) and (2.2), above, to separate the sense of 'can' in which he is interested from other senses. The unwanted senses of 'can' are the ones that associate naturally with truth—the ability sense of 'can' goes with 'can do', not with 'can be'. However, Cross' formalization forces a paraphrase involving 'can be true'; in (3.1), above, the function g is characterized in terms of whether "the truth of ϕ is within a's abilities."

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.

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

As far as I can see, the main value of such a step is not so much to correct any serious errors in the sentential formalism, but to provide a much more detailed and robust account of the selection function g.

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'. This doesn't seem right; 'If I tried I would' is a more intuitive conditional explication. This raises a fairly complex and delicate issue, one that is crucial for the logical analysis of ability.

The difference between Cross' semantics, which makes $\boldsymbol{Can}_a\phi$ true in case *some* outcome associated with a test of a's ability, and the idea that 'can' has to do with what *would* happen in such a trial, can be turned into an apparent counterexample. Suppose that my method of logging into John McCarthy's Stanford account is to telnet to cs.stanford.edu, enter 'jmc' at the login prompt, and then type in a random password. Since some (very improbable) outcome of applying this method will be successful, it seems to follow on Cross' semantics that

(5.1a) I can log into John McCarthy's Stanford account.

is true. That seems counterintuitive.

But this is a delicate matter. There are similar cases in which Cross' semantics is not obviously wrong. The California State Lottery may advertise

(5.1b) You can win over $15 Million in prize money this week!

Here, the intuition is that the advertisement is more misleading than straightforwardly false. And Cross' rule gives the right truth conditions if we want to classify this example as true but misleading.

The rule also appears to give the correct results for negations of ability statements. All the tests I can think of indicate that 'can't' is used to form the negation of sentences with positive 'can' as the auxiliary. For instance,

(5.2) I can't prove that theorem.

is the negation of Example (2.2a). According to Cross' rule, Example (2.0) is true if and only if every outcome of a normal attempt on my part to prove the theorem will be unsuccessful. If, however, we treated 'can' as analogous to Lewis' 'would', the semantic rule for 'can' would be

(Can2) $M \models_w <a>\phi$ if and only if for all $w' \in g(\phi, a, w)$, $M \models_{w'} \phi$.

According to this rule, Example (2.0) would amount to

(5.1) I might not prove that theorem if I tried to prove it.

which appears to be wrong. Note, for instance, that it is perfectly consistent to say

(5.2) Well, I might not manage to prove that theorem if I tried, but I believe I *can* prove it.

But if Example (5.1) were logically equivalent to Example (2.0), Example (5.2) would be Moore-paradoxical—it would have the appearance of inconsistency, rather than being a hedged statement of ability.

To sum up, the evidence concerning the modal status of 'can' appears to be mixed. This makes 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. There are two broad accounts of conditionals, which in [Thomason and Gupta, 1980] are called the *variably strict* and the *variably material* theories. According to the former sort of theory ([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 latter theory ([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. (Of course, maybe this opinion is prejudiced; but it is very hard to get around the fact that the negation of $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \psi$ is $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \neg\psi$, not $\boldsymbol{If}\ \phi\ \boldsymbol{then}\ \boldsymbol{Might}\,\neg\psi$.)

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 $g(\phi, a, w)$ 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 $M \models_{f(\phi,a,w)} \phi$.

The intuitive meaning of f is that $f(\phi, a, w)$ should be the closest world to w in which a tries to bring about ϕ.[5] 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

[5] 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.

that of **S4** or even **S5**, but I think it is unrewarding to press these details too far. Intuitions about $\boldsymbol{Can}_a\phi$ are not very robust for many sentences ϕ that can be formulated in English; and in fact if the aspectual type of ϕ is stative, a sentential modifier reading of 'can' tends to be forced in these cases. Thus, for instance,

(6.4a) Sam can be drunk.

is equivalent to

(6.4b) It can be true that Sam is drunk.

and

(6.5) Sam can not go home.

doesn't seem to make sense with narrow scope negation. You can force narrow scope by rephrasing (6.5) as

(6.6a) Sam can fail to go home.

but again, this is equivalent to

(6.6b) It can be true that Sam will not go home.

For these reasons, I do not take a formalization of the 'can' of occasional ability that treats it as an unrestricted sentential operator too seriously, though I believe that such a formalization can be instructive as a first approximation.

And for the same reasons, it is a little difficult to test the import of Rule (6.2) directly. Its plausibility rests mainly on intuitions that we canvassed before concerning the falsification of statements of ability; if that grizzly charges, and I try to climb the tree I chose and fail, this means that 'I can climb that tree' was false. To support this intuition, we need to use a unique world to determine the truth of the conditional.[6]

7. Ifs and cans

[Austin, 1956] inspired an important series of papers in the philosophical literature concerning ability. Austin enjoyed pointing out ways in which usage doesn't appear to fit the preconceptions of philosophers. Conditional constructions provide one such example. According to logical theories of 'if', including those of Austin's day and contemporary theories, conditional constructions express conditions of some sort in which the antecedent holds. In such theories, a case where ψ is vacuously qualified by ϕ—that is, a case in which $\phi \rightarrow \psi$ is logically equivalent to ψ—would be unusual. Also, $\phi \rightarrow \psi$ and $\neg\phi \rightarrow \psi$ jointly imply ψ, so it should be peculiar to say $\phi \rightarrow \psi$ when it is obvious that $\neg\phi \rightarrow \psi$ is true. [Austin, 1956] points out, however, that pairs of expressions like

[6] I realize that many readers will find it difficult to abandon the idea that many possibilites are involved in testing a statement of ability. Rule (6.2) is actually perfectly compatible with this intuition. Slack in the world that is appropriate for testing $\boldsymbol{Can}_a\phi$ can be represented by a multiplicity of choices for f. As I said, the issues here are the same as those that are involved in the semantics of the conditional; see [Stalnaker, 1980] for details and supporting argumentation.

(7.7a) I can answer that question.

and

(7.7b) I can answer that question if I try.

and like

(7.7c) I can answer that question if I choose.

are equivalent. Expressions like 'If I try' and 'If I choose' do not appear to add any genuine qualification to a conditional.

Austin also points out that Condition (7.7b) appears to imply (in some sense) Condition (7.7d), something that is certainly not the case with conditionals in general.

(7.7d) I can answer that question if I don't try.

Most of Austin's examples of such discepancies between received semantics and use have a connection (which Austin usually leaves implicit) to issues of philosophical importance. In this case, the philosophical issue is the freedom of the will, and the connection is that if the 'can' of ability ultimately has some sort of conditional import, then it may be possible to maintain that at there are always many incompatible things that an agent can do—even in a deterministic universe.

I will not discuss the connection to the issue of freedom. As the reasons for thinking that we inhabit a deterministic universe have been undermined, it seems far less important to make a case for compatibilism than to make sense of the consequences for common sense and scientific thinking of the fact that things are thoroughly nondeterministic. Also, I agree with Austin that the more eager we are to understand the perennial issues of philosophy the less likely we are to have the resources we need to discuss these issues in a way that can shed new light on them.

But I think that we may be in a position to explain how the phenomena that Austin noted do not show that the interaction of conditionals with 'can' does not provide evidence for a nonstandard interpretation of 'if ... then'.

According to the theory we have presented,

(7.8a) $\boldsymbol{Can}_a \phi$

is equivalent to

(7.8b) ***If*** a tries to bring about ϕ ***then*** ϕ.

Then Condition (7.7b) is equivalent to

(7.9a) ***If*** I try to bring about [I answer that question] ***then***
if I try to bring about [I answer that question] ***then***
I answer that question.

But in just about any logic of conditionals, Condition (7.9a) is equivalent to

(7.9b) ***If*** I try to bring about [I answer that question] ***then***
I answer that question.

So Austin's first alleged anomaly is not an anomaly at all, but is explained by the law of contraction for the antecedents of conditionals.

Austin's second alleged anomaly is illustrated by the equivalence of Condition (7.7b) with Condition (7.7d). Now, Condition (7.7b) amounts to Condition (7.9a). And Condition (7.7d) is equivalent to

(7.9) ***If*** ¬[I try to bring about [I answer that question]] ***then***
if I try to bring about [I answer that question] ***then***
I answer that question.

Now, this is certainly not a logical equivalence, but it follows from a plausible enough constraint on the selection function for conditionals: that the value this function delivers for antecedent 'I try to bring about [I answer that question]' and world w is the same as the value it delivers for antecedent 'I try to bring about [I answer that question]' and the world w' that is selected for antecedent '¬[I try to bring about [I answer that question]]' and world w. Essentially, this condition expresses the conditional independence of whether a trial is made and the result of the trial. Although we can't in general expect this condition to hold for all antecedents—for instance, observation actions in which an attempt to perform the observation will affect the results are problematic—we can hope that it applies at least to everyday actions.

8. Providing 'can' with actions as arguments

In Section 4 I alluded to the evidence against a formalism that treats the clausal argument of the 'can' of occasional ability as sentential. The most straightforward way to correct this implausibility is to divide the work of formalizing natural language sentences 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 logical forms of natural language sentences (in this case, of the sentences of English) that makes a representation of an appropriate individual action available for clauses that are appropriate arguments of 'can'.

Using this division of labor, I can appeal to work in eventuality-oriented natural language semantics, such as [Parsons, 1990, Higginbotham *et al.*, 2000] for part (ii), and can confine myself to (i), the purely logical component of the project.

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) methods of performing higher-level actions in performing sequences of 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.

We will take as our start the Situation Calculus.[7] This too could be considered to be a simplification; but the Situation Calculus has proved to be remarkably robust. I adopt a

[7] [McCarthy and Hayes, 1969, Shanahan, 1997].

formalism that allows knowledge-how to be expressed, 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].

9. 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 here and in the present draft) 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.

(9.1) $\forall x_s[\textsc{Pre}(A, x) \rightarrow \textsc{Post}(\textsc{Result}(A, x))]$.

Here, PRE is the precondition for A and POST is the postcondition or effect of a.[8]

10. Situation Calculus with Explicit Ability

Often, a causal axiom of the form Condition (9.1) 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 in more cases where 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, common sense 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

[8] To simplify things, I am writing $\textsc{Pre}(A, s)$ rather than $\textsc{Holds}(\textsc{Pre}(A), s)$. 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.

(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 try to talk to her with this knowledge, or, more typically, I may not know that I can talk to her but will try talk to her simply hoping that I can do so, and without any clear fallback plan in case of failure.[9]

Here is a relatively simple way to formalize these ideas in a Situation Calculus format. First, we assume that certain actions (generally, these will correspond to the ones that are expressed by telic verbs in natural language) have causal axioms like Condition (9.1) associated with them. For instance, such an axiom could connect telephoning my wife with my having a verbal communication channel open to my wife.

Second, we need to deal with the fact that trying to perform an action can be different from performing it. As a working start, I will postulate a function $\mathbf{f}$ taking actions to actions. This function inputs an action $\mathbf{A}$ and situation $\mathbf{s}$, and outputs the agent's standard method of attempting to perform $\mathbf{A}$ in $\mathbf{s}$. I am making a number of simplifications here; let me explain them briefly. (1) An agent may have many ways of performing $\mathbf{A}$ in $\mathbf{s}$. I am assuming that one of these is the one that the agent *would* use if it tried to perform $\mathbf{A}$. So $\mathbf{f}$ has a certain amount of counterfactual content. (2) An agent may have no standard way of performing $\mathbf{A}$ in $\mathbf{s}$; also, whatever "having a standard way of performing $\mathbf{A}$ in $\mathbf{s}$" means, it should entail that the agent knows how to $\mathbf{A}$ in $\mathbf{s}$. In fact, knowing how and ability are intimately connected. 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 $\mathbf{f}(\mathbf{A}, \mathbf{s}) = \text{NULL}$ means that $\mathbf{f}$ is undefined for $\mathbf{A}$ in $\mathbf{s}$, and (ii) assuming that $\mathbf{f}(\mathbf{A}, \mathbf{s}) = \text{NULL}$ when the agent doesn't know how to $\mathbf{A}$ in $\mathbf{s}$.

This formalization automatically generates a regress of actions $\mathbf{A}, \mathbf{f}(\mathbf{A}, \mathbf{s}), \mathbf{f}(\mathbf{f}(\mathbf{A}, \mathbf{s}), \mathbf{s}), \ldots$. $\mathbf{A}$ is, say, the initial action that our fixed agent wants to perform. $\mathbf{f}(\mathbf{A}, \mathbf{s})$ is the action that the agent needs to perform in order to try to perform $\mathbf{A}$. But to perform $\mathbf{f}(\mathbf{A}, \mathbf{s})$, the agent has to try to perform it. $\mathbf{f}(\mathbf{f}(\mathbf{A}, \mathbf{s}), \mathbf{s})$ is the agent's standard way of trying to perform $\mathbf{f}(\mathbf{A}, \mathbf{s})$. And so forth.

Some notation and terminology.

> $\mathbf{f}^0(\mathbf{A}, \mathbf{s}) = \mathbf{A}; \mathbf{f}^{n+1}(\mathbf{A}, \mathbf{s}) = \mathbf{f}(\mathbf{f}^n(\mathbf{A}, \mathbf{s}), \mathbf{s})$.
>
> $\mathbf{A}_n^s = \mathbf{f}^n(\mathbf{A}, \mathbf{s})$.
>
> The $\mathbf{s}$-*order* of an action $\mathbf{A}$ is the least n such that $\mathbf{f}^n(\mathbf{A}, \mathbf{s}) = \mathbf{f}^{n+1}(\mathbf{A}, \mathbf{s})$.

So if $\mathbf{A}$ is stacking Block 1 on Block 2, and both blocks are clear, then maybe

(1) $\mathbf{A}_1^s$ is grasping Block 1 and then moving it to Block 2,

(2) $\mathbf{A}_2^s$ is clearing Gripper 1, moving it to Block 1, then gripping Block 1, then moving Gripper 1 to Block 2, then releasing Block 1,

[9] Typically, people will act on common sense plans without anything like a proof that they will succeed. All that is seems to be required is the absense of a proof that they will fail and relatively low risk or cost of trying to achieve the goal, in relation to the benefits of the goal.

(3) $\mathbf{A}_3^s$ is moving Gripper 1 to the table, then releasing the block in Gripper 1, then moving Gripper 2 to Block 1, then then gripping Block 1, then moving Gripper 1 to Block 2, then releasing Block 1.

In this robotics-flavored example, this process can be taken pretty far, and how far you want to take it will depend on how much of the mechanical part of the process you want to take into account. In human examples, how far you can take it and at what point you stop talking about actions becomes a philosophical problem.

People familiar with hierarchical planning formalisms will recognize the close similarity of this regression to action decomposition.

If the regression for $\mathbf{A}$ continues infinitely, $\mathbf{A}$ will, it seems, not be performable by the agent. The intuition behind this does not depend on the impossibility of infinite series in nature, but on the idea that a feasible plan has to be grounded in feasible actions, where part of what is meant here by "feasible" is that the performance of these grounding actions are not constituted by the performance of other actions. Both the process of planning and the execution of plans seem to be incompatible with regresses of this kind. Therefore, we can place the following condition on $\mathbf{f}$.

(10.1) For all actions $\mathbf{A}$, there is an n such that $f^n(\mathbf{A}) = f^{n+1}(\mathbf{A})$.

Condition (10.1) entails the existence of actions $\mathbf{A}$ such that $\mathbf{f}(\mathbf{A}) = \mathbf{A}$. Actions that are fixpoints of $\mathbf{f}$ resemble in some ways the so-called "basic actions" that have been discussed in the literature on the philosophy of action.[10]

We denote the function $\mathbf{f}$ by a function letter TRY of the planning formalism. With the incorporation into the formalism of a distinction between an action $\mathbf{A}$ and the action of trying to do $\mathbf{A}$, 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.

(10.2) $\forall x_s[\textsc{Can}(A, x) \to \textsc{Post}(\textsc{Result}(\textsc{try}(A, s), x))]$.

Such axioms correspond to platitudes of the sort

(10.3) 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 an action $\mathbf{A}$ of $\mathbf{s}$-order 0, where s denotes $\mathbf{s}$. Then the success conditions for $\mathbf{A}$, i.e. the conditions under which trying to perform $\mathbf{A}$ (in this case simply performing $\mathbf{A}$) will achieve the effects conventionally associated with A are simply the preconditions of $\mathbf{A}$. Here, the definition of $\textsc{Can}(A, s)$ is simply this, where $\mathbf{A}$ denotes a 0-order action.

(10.4) $\textsc{Can}(A, s)$ amounts to $\textsc{Pre}(A, s)$.

[10] For background in the philosophy of action, see [Brand, 1984, Care and Landesman, 1968, Danto, 1973, Goldman, 1970].

But, as we have noticed, trying to perform an action may lead us to try to perform other actions—for instance, trying to open a certain door can involve trying to turn the doorknob. Suppose, then, that A_0 is a 1-order action. For instance, A_0 could be opening the door, and $\textsc{try}(A, s)$ (i.e., A_1^s) could be turning the doorknob and then pushing the door. Say that $\textsc{Pre}(A_0^s, s)$ is 'The door isn't stuck' and $\textsc{Pre}(A_1^s, s)$ is 'The door isn't locked'.[11] Then, in this case, $\textsc{Can}(A_0, s)$ amounts to 'The door is neither stuck nor locked'.

The general reduction of $\textsc{Can}$, then, for a constant A_0^s denoting an n-order act, is this.

(10.5) $\textsc{Can}(A_0^s, s)$ amounts to $\textsc{Pre}(A_0^s, s) \wedge \ldots \wedge \textsc{Pre}(A_n^s, s)$.

The account of the nongeneric 'can' of ability that has emerged from this simple treatment in the Situation Calculus is very similar to the more abstract, conditional account of Section 6, although the difference between the two formalisms makes a direct comparison difficult. 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 Situation Calculus treatment is that it provides a method of representing knowledge about actions 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, we can recover useful conditions.

11. Contextual Variation in 'Can'

By focusing on the nongeneric 'can' of ability, I have eliminated most of the interactions between context and 'can' that have been discussed in the literature. For instance, [Kratzer, 1977], one of the best-known studies of contextual effects on 'can', is concerned only with contextual effects of background information on the sentence modifier 'can'—the 'can' of possibility. The 'can' of occasional ability is not affected by, for instance, a contextual set of possibilities that we take to be supposed or believed; it is bound to the set of genuine practical, historical possibilities.

However, the 'can' of occasional ability is not entirely decontextualized; it is, in fact, subject to a sort of contextual variation that is centrally important to practical reasoning, and that is in some ways more interesting than the contextual effects that have been discussed in the literature.

These effects are illustrated by the following examples.

(11.1a) I can't meet you this afternoon, because I have to teach a class.

(11.1b) I can't drive today, because I forgot my driver's license.

(11.1c) I can't eat that candy, because I'm allergic to peanuts.

It is easy to imagine circumstances in which Condition (11.1a–c) are natural and plausible things to say, although they would be false according to the conditional account of 'can'. For instance, Condition (11.1a) certainly doesn't mean that if I tried to meet you this afternoon I wouldn't meet you. The suggestion in each case is that if I tried I would succeed narrowly

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

in performing the designated action, but the performance would violate some commitment or policy in each case.[12]

This appears to provide linguistic confirmation of the idea of [Bratman, 1987, Bratman *et al.*, 1988], that commitments to plans constrain practical deliberation by acting as a filter that eliminates new plans that conflict with the existing ones.

Model-theoretically, we can represent the fundamental practical constraint from Section 10 that actions can only be performed when the agent is able to perform them as determinining a set $\mathcal{H}$ of situation calculus histories, where a history rooted in an initial situation $\mathbf{s}$ is simply a sequence $\mathbf{h}$ (which I will suppose to be infinite) of actions. The fundamental constraint amounts to this: *all histories rooted in* $\mathbf{s}$ *must be feasible for* $\mathbf{s}$, where $\mathbf{h} = \langle \mathbf{A}_1, \mathbf{A}_2, \ldots \rangle$ is feasible for $\mathbf{s}$ if for all $i \geq 0$, $\textsc{Can}(A_{i+i}, s_i)$ holds, where A denotes $\mathbf{A}$ and s_i denotes $\mathbf{s}_i$. Here, $\mathbf{s}_0 = \mathbf{s}$ and for all $i > 0$, $\mathbf{s}_{i+1} = \mathbf{r}(\mathbf{A}_{i+1}\mathbf{s}_i)$.

Bearing in mind the reduction Condition (10.5) of a formula having the form $\textsc{Can}(A, s)$ to a conjunction of preconditions, the fundamental constraint simply restricts the possible histories to those in which the preconditions of every action that is performed are met, as well as the preconditions of all its constitutive enabling actions. We want a *planning context* to further restrict the possible histories, to those that are compatible with the plans to which the agent is already committed.

So a planning context for a situation $\mathbf{s}$ will simply be a subset $\mathcal{H}$ of the set of feasible histories rooted in $\mathbf{s}$. In this draft, I have not tried to explicitly develop a language in which complex plans can be expressed. But I am assuming a fairly flexible domain of plans, including things like "at some point in the future before tomorrow, get out the hose, then attach it, then water the lawn, then put the hose away." The idea is that the adoption of any such plan creates a new planning context by removing from the current context $\mathcal{H}$ any histories that are incompatible with it.

A formula $\textsc{Can}(A, s)$ is true relative to a planning context $\mathcal{H}$ for $\mathbf{s}$ if there is a sequence $\mathbf{h} \in \mathcal{H}$ such that $\mathbf{A}$ is the first action in the sequence $\mathbf{h}$. This means that not only are all the preconditions of $\mathbf{A}$ met in $\mathbf{s}$, as well as all the preconditions of $\mathbf{A}$'s constitutive enabling actions, but that the immediate performance of $\mathbf{A}$ is not incompatible with the plans that the agent has formed so far.

Afterthought to this section, added October, 2005. Everything in this section was implicitly relativized to a single agent a. Differences between $\textsc{Can}(a, A, s)$ and $\textsc{Can}(a', A, s)$ would emerge, on this account, from differences between the preconditions that apply to different agents. An agent that is at a door satisfies the one of the preconditions for opening the door; an agent that is not at the door will not satisfy this precondition. In this way, the account preserves Cross' insight that 'can' is indexed to agents. There is another kind of generic construction involving 'can' that shows up in agentless passive, as in 'That watch can be fixed', which I think means something like this. "Grant me a properly qualified agent, and normal circumstances. Then under these circumstances, the agent can fix the watch." This construction has a possibilistic flavor.

[12] The 'can't' of Condition (11.1c), for instance, doesn't seem to signal merely that the consequences of eating the candy would be harmful. Note that it is peculiar to say 'I can't touch that skillet; I'd burn myself'.

12. Conclusion

In this draft, I have only tried to formalize the simplest version of an action-and-change formalism capable of explicitly representing ability.

There actually is a great deal of work to be done here, and 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.

Besides these extensions, there are a number of simplifications in the presentation that need to be done away with. In particular, there is a mismatch between my examples of actions illustrating the planning formalism presented in Section 10 and the language of the formalism. I use action descriptions like 'Turn the doorknob, then push the door', but the planning language itself does not provide for terms denoting action sequences. A more adequate planning language would include a family of action-term operators including perhaps sequencing, conditional actions, and Kleene star. This step would begin to close the gap between the planning formalism and dynamic logics.

References

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

[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.

[Bratman *et al.*, 1988] Michael E. Bratman, David Israel, and Martha Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[Bratman, 1987] Michael E. Bratman. *Intentions, Plans and Practical Reason.* Harvard University Press, 1987.

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

[Carlson and Pelletier, 1995] Greg N. Carlson and Francis Jeffrey Pelletier, editors. *The Generic Book.* Chicago University Press, Chicago, Illinois, 1995.

[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.

[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.

[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.

[McCarthy and Hayes, 1969] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[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.

[Pollack and Horty, 1999] Martha Pollack and John F. Horty. There's more to life than making plans. *The AI Magazine*, 20(4):71–84, 1999.

[Shanahan, 1997] Murray Shanahan. *Solving the Frame Problem*. The MIT Press, Cambridge, Massachusetts, 1997.

[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.

[Stalnaker, 1980] Robert C. Stalnaker. A defense of conditional excluded middle. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pages 87–104. D. Reidel Publishing Co., Dordrecht, 1980.

[Thomason and Gupta, 1980] Richmond Thomason and Anil Gupta. A theory of conditionals in the context of branching time. *The Philosophical Review*, 80:65–90, 1980.

[Thomason and Stalnaker, 1973] Richmond H. Thomason and Robert C. Stalnaker. A semantic theory of adverbs. *Linguistic Inquiry*, 4:195–220, 1973.

[Thomason, 1970] Richmond H. Thomason. Indeterminist time and truth-value gaps. *Theoria*, 36:246–281, 1970.