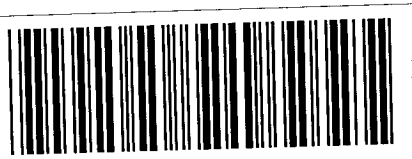

Call Number:

Location:

Maxcost: rec

DateReq: 9/8/2004 ☐ Yes

Date Rec: 9/9/2004 ☐ No

☐ Conditional

Borrower: IRU

LenderString: TXU,*OIT,PGP,IXA,NHM

ILL: 8337677

Title: Computational intelligence = Intelligence informatique.

Author:

Edition:

Imprint: Ottawa : National Research Council of Canada = Conseil national

Article: M.E.Bratman, D.J. Israel, and M.E. Pollack "Plans and resource- bounded practical reasoning"

Vol: 4 **No.:** 4 **Pages:** 14-23 **Date:** 1988

Borrowing Notes: ClioID: f0023296////

Fax:

ILL: 8337677 :Borrower: IRU :ReqDate: 20040908 :NeedBefore: 20041007
:Status: IN PROCESS 20040909 :RecDate: :RenewalReq:
:OCLC: 12073389 :Source: Clio :DueDate: :NewDueDate:
:Lender: TXU,*OIT,PGP,IXA,NHM
:CALLNO: *Lender's OCLC LDR: v.4- 1988- :TITLE: Computational intelligence = Intelligence informatique. :IMPRINT: Ottawa : National Research Council of Canada = Conseil national de recherches du Canada, 1985- :ARTICLE: M.E.Bratman, D.J. Israel, and M.E. Pollack "Plans and resource- bounded practical reasoning" :VOL: 4 :NO: 4 :DATE: 1988 :PAGES: 14-23 :VERIFIED: OCLC ISSN: 0824-7935 [Format: Serial] :PATRON: Tran, Cao Son :SHIP TO: New Mexico State University/1305 Frenger Mall/Interlibrary Loan, Zuhl Library/Box 30003 Dept 3475/Las Cruces NM USA 88003 :BILL TO: same, FEIN: 85-6000401 :SHIP VIA: TAE/ARIEL 128.123.193.167 :MAXCOST: rec :COPYRT COMPLIANCE: CCL :FAX: (505) 646-4335 ARIEL Address: 128.123.193.167 :E-MAIL: ill@lib.nmsu.edu :BORROWING NOTES: ClioID: f0023296//// :AFFILIATION: @/am-BCR LVIS IFM CARLA TAE Reciprocal agreements welcome. :LENDING CHARGES: :SHIPPED: :SHIP INSURANCE: :LENDING RESTRICTIONS: :LENDING NOTES: :RETURN TO: :RETURN VIA:

ShipVia: TAE/ARIEL 12

NeedBy: 10/7/2004

Return To:

Oregon Institute of Technology
Library
3201 Campus Drive
Klamath Falls, OR 97601-8801

Ship To:

New Mexico State University
1305 Frenger Mall
Interlibrary Loan, Zuhl Library
Box 30003 Dept 3475
Las Cruces NM USA 88003

ILL: 8337677 **Borrower:** IRU

Req Date: 9/8/2004 **OCLC #:** 12073389

Patron: Tran, Cao Son

Author:

Title: Computational intelligence = Intelligence i

Article: M.E.Bratman, D.J. Israel, and M.E. Pollack "Plans and resource- bounded practical reasoning"

Vol.: 4 **No.:** 4

Date: 1988 **Pages:** 14-23

Verified: OCLC ISSN: 0824-7935 [Format: Serial]

Maxcost: rec **Due Date:**

Lending Notes:

Bor Notes: ClioID: f0023296////

Plans and resource-bounded practical reasoning

MICHAEL E. BRATMAN
Department of Philosophy and Center for the Study of Language and Information, Stanford University, Stanford, CA 94305, U.S.A.

AND

DAVID J. ISRAEL AND MARTHA E. POLLACK
Artificial Intelligence Center and Center for the Study of Language and Information, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, U.S.A.

Received September 13, 1987

Revision accepted September 19, 1988

An architecture for a rational agent must allow for means-end reasoning, for the weighing of competing alternatives, and for interactions betwen these two forms of reasoning. Such an architecture must also address the problem of resource boundedness. We sketch a solution of the first problem that points the way to a solution of the second. In particular, we present a high-level specification of the practical-reasoning component of an architecture for a resource-bounded rational agent. In this architecture, a major role of the agent's plans is to constrain the amount of further practical reasoning she must perform.

Key words: planning, practical reasoning, resource bounds.

L'architecture d'un agent rationnel doit permettre le raisonnement procédant des fins aux moyens, le choix entre différentes actions possibles, et l'interaction entre ces deux modes de raisonnement. Elle doit aussi tenir compte des conséquences des limites de ressources disponibles. Nous esquissons ici une solution au premier problème qui indique comment on pourrait résoudre le second. Nous proposons, en particulier, une spécification abstraite d'un module de génération de plans pour un agent rationnel dont les ressources sont bornées. Dans cette architecture, le rôle principal des plans d'un agent est de limiter les ressources devant être consacrés au raisonnement.

Mots clés : planification, raisonnement pratique, limites de ressources.

Comput. Intell. **4**, 349–355 (1988)

1. Introduction

Rational behavior — the production of actions that further the goals of an agent, based upon her conception of the world — has long interested researchers in artificial intelligence, who are attempting to build machines that behave rationally, as well as philosophers of mind and action, decision theorists, and others who are attempting to provide an account of human rationality. Each of these research traditions has tended to concern itself with a different facet of the problem.

Within AI, much attention has been given to the "planning problem," namely, the problem of automating means-end reasoning. AI solutions to the planning problem generally consist of methods for searching the space of possible actions to compute some sequence of actions that will achieve a particular goal or conjunction of goals. Work in this area has resulted in a number of extremely useful techniques for representing and reasoning about actions and their effects (Fikes and Nilsson 1971; Sacerdoti 1977; Georgeff and Lansky 1987*b*; Georgeff 1987*b*).

Within decision theory (DiFinetti 1975; Jeffrey 1983; Savage 1972), the primary concern has been somewhat different: competing alternatives are taken as given, and the problem is to weigh these alternatives and decide on one of them. A completed means-end analysis is implicit in the specification of the competing alternatives.

It is clear that rational agents must both perform means-end reasoning and weigh alternative courses of action; so an adequate architecture of intelligent artificial agents must therefore include capabilities for both. The design of such an architecture must also specify how these capacities interact. But there is yet another problem. All this must be done in a way that recognizes the fact that agents, whether humans or robots, are *resource bounded*: they are unable to perform arbitrarily large computations in constant time.[1] To what extent have the AI and decision-theoretic traditions faced up to questions raised by the phenomenon of resource boundedness?

In decision-theoretic accounts, an agent is seen as selecting a course of action on the basis of her subjective expected utility, which is a function of the agent's beliefs and desires. For an idealized, resource-*unbounded* agent, this may be a plausible model: perhaps such an agent could, at each instant of time, compute which course of action currently available would maximize its expected utility. But, of course, for real agents it takes time to do such computations — and the more complicated they are, the more time it takes. This is a problem because the more time spent on deliberation, the more chance there is that the world will change in important ways — ways that will undermine the very assumptions on which the deliberation is proceeding.

What about AI planning systems? With some exceptions (Albus 1981; Georgeff and Lansky 1987*a*; Kaelbling 1987; Schmidt 1985; Durfee and Lesser 1986), these have typically been designed to construct plans prior to, and distinct from, their execution. It is recognized that the construction of plans takes time. However, these plans have been constructed for a set of future conditions that are known in advance and are frozen. The implicit assumption is that the conditions for which a plan is being formed, the so-called start state, will not change prior to execution. And when it is assumed that the plans will be executed in single-agent environments, in which the only state changes are a result of the single agent's actions, there is no concern that the world will change in unexpected

[1]Problems of resource boundedness have been forcefully pointed out by Herbert Simon; see, for example, Simon (1957).

ways during execution. But, of course, the world does not actually stay fixed during an indefinitely long planning period. Nor do true single-agent environments exist: even if the environment contains no other human or robot agents, nature often intrudes.[2]

During the time it takes to engage in practical reasoning, the world can change in important ways. This fact poses the problem of resource boundednes that concerns us here.

So we have two problems. First, an architecture for a rational agent must allow for means-end analysis, for the weighing of competing alternative, and for interactions between these two forms of reasoning. Second, this architecture must address the problem of resource boundedness. We sketch a solution of the first problem that points the way to a solution of the second. In particular, we present a high-level specification of the practical-reasoning component of an architecture for a resource-bounded rational agent. In this architecture, a major role of the agent's plans is to constrain the amount of further practical reasoning she must do.

2. The functional roles of plans

Figure 1 is a block diagram of an architecture for practical reasoning in resource-bounded agents. It can be classified as a belief/desire/intention (BDI)-architecture: it includes fairly direct representations of the agent's beliefs, desires, and intentions. We view the agent's intentions as structured into larger plans. We distinguish between the plans that the agent has actually adopted, which are represented in Fig. 1 in the oval labeled "Intentions structured into plans," and the plans-as-recipes, or operators, that the agent knows about, which are represented in the oval labeled "Plan library." The plan library might be seen as a subset of the agent's beliefs: specifically, her beliefs about what actions would be useful for achieving which effects under specified conditions. We shall reserve the term "plan" to refer to those plans an agent has actually adopted.

In addition to the information stores, which are denoted by ovals in the figure, there are a number of processes, denoted by rectangles. Our concern will be with four of these: the *means-end reasoner*, the *opportunity analyzer*, the *filtering process*, and the *deliberation process*. Together these constitute a practical-reasoning system, i.e., a system by which an agent forms, fills in, revises, and executes plans.

Underlying the architecture depicted in Fig. 1 is an account of the functional roles of an agent's plans not just in producing action, but also in constraining futher, practical reasoning — an account so far developed largely by Bratman (1987). In this account, an agent's existing plans make practical reasoning more tractable in two ways: as input to the means-end reasoner, they provide a clear, concrete purpose for reasoning; and as input to the filtering process, they narrow the scope of deliberation to a limited set of options. We shall briefly describe this conception and then explain how it is realized in the architecture depicted in Fig. 1.

The fundamental observation of our approach is that a rational agent is committed to doing what she plans.[3] The nature of this commitment is quite complex (Bratman 1987), but involves at least certain characteristic roles in further practical reasoning.[4] For example, once an agent has formed a plan to attend a particular meeting at 1:00, she need not continually weigh the situation at hand in a wholly unfocused manner. Instead, she should reason about how to get there by 1:00; she need not consider options incompatible with her getting there by 1:00; and she can typically ground her further reasoning on the assumption that she will indeed attend the meeting at 1:00. This example illustrates three roles that an agent's plans will play in her further reasoning: they will drive means-end reasoning, they will provide constraints on what options need be seriously considered, and they will influence the beliefs on which further practical reasoning will be based. In this paper, we focus primarily on the first two roles.

Consider the constraining role of plans. Other things being equal, an agent's plans should be consistent, both internally and with her beliefs. Roughly speaking, it should be possible for her plans, taken together, to be executed successfully in a world in which her beliefs are true. As a result of this demand for consistency, options that are inconsistent with her existing plans and beliefs will be filtered out.

Of course, prior plans may be subject to reconsideration or abandonment in light of changes in belief. But if an agent constantly reconsiders her plans, they will not limit her deliberation in the way they need to for a resource-bounded agent. This means that an agent's plans should be reasonably stable, i.e., they should be relatively resistant to reconsideration and abandonment.[5]

Given the requirement of stability, plans should also be partial. In addition to bounded computational resources, agents have bounded knowledge. They are neither prescient nor omniscient: the world may change around them in ways they are not in a position to anticipate. Hence highly detailed plans about the far future will often be of little use, the details not worth bothering about.

Plans can be partial in at least two different ways. They may be *temporally partial*, accounting for some periods of time and not for others. An agent may plan to give a lecture from 10:00 until noon, to pick up a book at the bookstore on the way back from the lecture, to attend a meeting from 1:00 to 2:30, and to pick up her child at school at 4:00; she may not yet have decided what to do between 2:30 and the time she leaves for her child's school.

More important for our purposes is the potential for *structural partiality* in plans. Agents frequently decide upon ends, leaving open for later deliberation questions about means to

[2]Even when it is assumed that the world changes only as a result of the agent's actions, it is still infeasible for that agent to consider all possibilities ahead of time. In consequence, the primary capabilities of many practical planning systems were augmented to allow for the monitoring of plan execution and for replanning (Fikes and Nilsson 1971; Sridharan and Bresina 1982; Wilkins 1984). However, the replanning modules that were built had much the same character as the planning modules themselves: they operated under the assumption that the world around them was frozen during replanning. Recently, there has been a growing concern with developing representations for multi-agent domains (Georgeff 1987*a*; Lansky 1987; McDermott 1985).

[3]The reader should recall our distinction between intended plans and plans-as-recipes.

[4]An attempt at constructing a formal model for commitment is made by Cohen and Levesque (in press).

[5]We discuss these matters further in Sect. 5.

those ends.[6] An agent may, for example, first decide to pick up a book at the bookstore, postponing decisions about what route to take to get there and whether to use Visa or MasterCard to pay. The structural partiality of plans is the reason we speak of their decomposition into intentions: for example, we shall speak of an agent's "filling in" her plan to buy a book with an intention to pay for it with her MasterCard. We shall also be concerned with the interaction between a plan's decomposition and the requirement of consistency. A plan to spend all of one's cash at lunch is inconsistent with a plan to buy a book that includes and intention by pay for it with cash, but is not necessarily inconsistent with a partial plan merely to purchase a book, since the book may be paid for with a credit card.

The characteristic process of means-end reasoning suggests the second way in which plans focus the practical-reasoning process. Plans, while potentially partial, must be *means-end coherent*: as time goes by, they must be filled in with subplans that are at least as extensive as the agent believes necessary to execute the plan successfully.[7] As a result of the demand for means-end coherence, prior, partial plans can be seen to pose problems for further practical reasoning. Once the agent has decided to read a certain book today, a means-end problem is posed: how will she get the book? Will she go to the library to borrow a copy of it, or will she stop by the bookstore and purchase one? Once she has formed an intention to read the book, her reasoning can focus on deciding how to do so, rather than on assessing all the options that are currently available.

3. The larger architecture

We can now return to the architecture illustrated in Fig. 1. Let us assume, for expository purposes, an agent who embodies this architecture and who has already adopted some structurally partial plans, and let us consider the practical reasoning she will perform. Her means-end reasoner will be invoked for each of her existing partial plans, to propose subplans that complete it. Means-end reasoning may occur at any time up to the point at which a plan is in danger of becoming means-end incoherent; at that point it must occur, proposing options that may serve as subplans for the plan in question. The means-end reasoner may propose a number of options, all of which are means to a particular end: for example, it may propose going to the bookstore and going to the library as alternative means to getting a desired book.

Not all options are proposed as a result of means-end reasoning. Changes in the agent's environment may lead to changes in her beliefs, which in turn may result in her considering new options that are not means to any already intended end. The opportunity analyzer in Fig. 1 is the component that proposes options in response to perceived changes in the environment. Such opportunities may be welcome or unwelcome. Some changes may lead to previously unexpected opportunities for satisfying desires; others to opportunities for avoiding unexpected threats.

Once options have been proposed, either by the means-end reasoner or by the opportunity analyzer, they are subject to filtering. So far, we have suggested how one of the components of the filtering process, the compatibility filter, operates. (In Sect. 4, we explain how the behavior of the overall filtering process is affected by the other component, the filter override mechanism.) The compatibility filter checks options to determine compatibility with the agent's existing plans. Options deemed compatible are *surviving options*. Surviving options are passed along to the deliberation process and, when there are competing surviving options, they are weighed against one another. The deliberation process produces intentions, which are incorporated into the agent's plans.

It is essential that the filtering process be computationally efficient relative to deliberation itself. After all, the motivation for introducing this process into the architecture was to reduce the amount of computation in practical reasoning. Here a variety of ideas can be explored. One is to delimit types of incompatibility that can be checked in a computationally tractable manner. Thus, for example, one might define a measure of spatiotemporal separation between options and design the compatibility filter so that it rules out all and only those options that overlap inappropriately with already intended actions. For an important class of cases, such as a scheme can be implemented as a polynomial-time constraint-propagation algorithm over intervals (Kautz and Vilain 1986). Another idea would be to employ a tractable system of defeasible reasoning involving imperfect, albeit still useful, filters. Such filters may be "leaky," in that they sometimes let through options that are in fact incompatible, or they may be "clogged," in that they sometimes block options that are in fact compatible.

What happens when the agent comes to believe that a prior plan of hers is no longer achievable? A full development of this architecture would have to give an account of the ways in which a resource-bounded agent would monitor her prior plans in the light of changes in belief. However this is developed, there will of course be times when an agent will have to give up a prior plan in light of a new belief that this plan is no longer executable. When this happens, a new process of deliberation may be triggered indirectly in one of two ways. First, the abondoned plan may be the specification of the agent's means to some intended end. In this case, the agent's larger plan will be threatened with means-end incoherence, which would normally trigger reasoning of the sort we have already described. But sometimes the abandoned plan may not be the specification of a means to a presently intended end. Still, we can suppose that this plan was initially adopted as a way of satisfying some desire. If the agent sitll has this desire, it may lead to further deliberation should an appropriate opportunity arise.

4. Filtering and overriding

The account of practical reasoning given so far is incomplete in an important way. Recall that agents are not only resource-bounded, but also knowledge-bounded. So a rational agent's current plans must not have irrevocable control over her future deliberation and behavior. Rather, a rational agent should sometimes be willing to reconsider her plans in light of unan-

[6]Hence, structural partiality is related to the partiality of plans produced by traditional, hierarchical planners, such as NOAH (Sacredoti 1977). However, whereas these planners used partial plans only as intermediate representations in the plan formation process, we are suggesting the usefulness of acting on the basis of partial plans. PRS (Georgeff and Lansky 1987*a*) is an example of a system that makes use of structurally partial plans during execution.

[7]Further development of this architecture requires the construction of techniques for detecting threats to means-ends coherence, techniques that are compatible with the demands of relative computational efficiency.

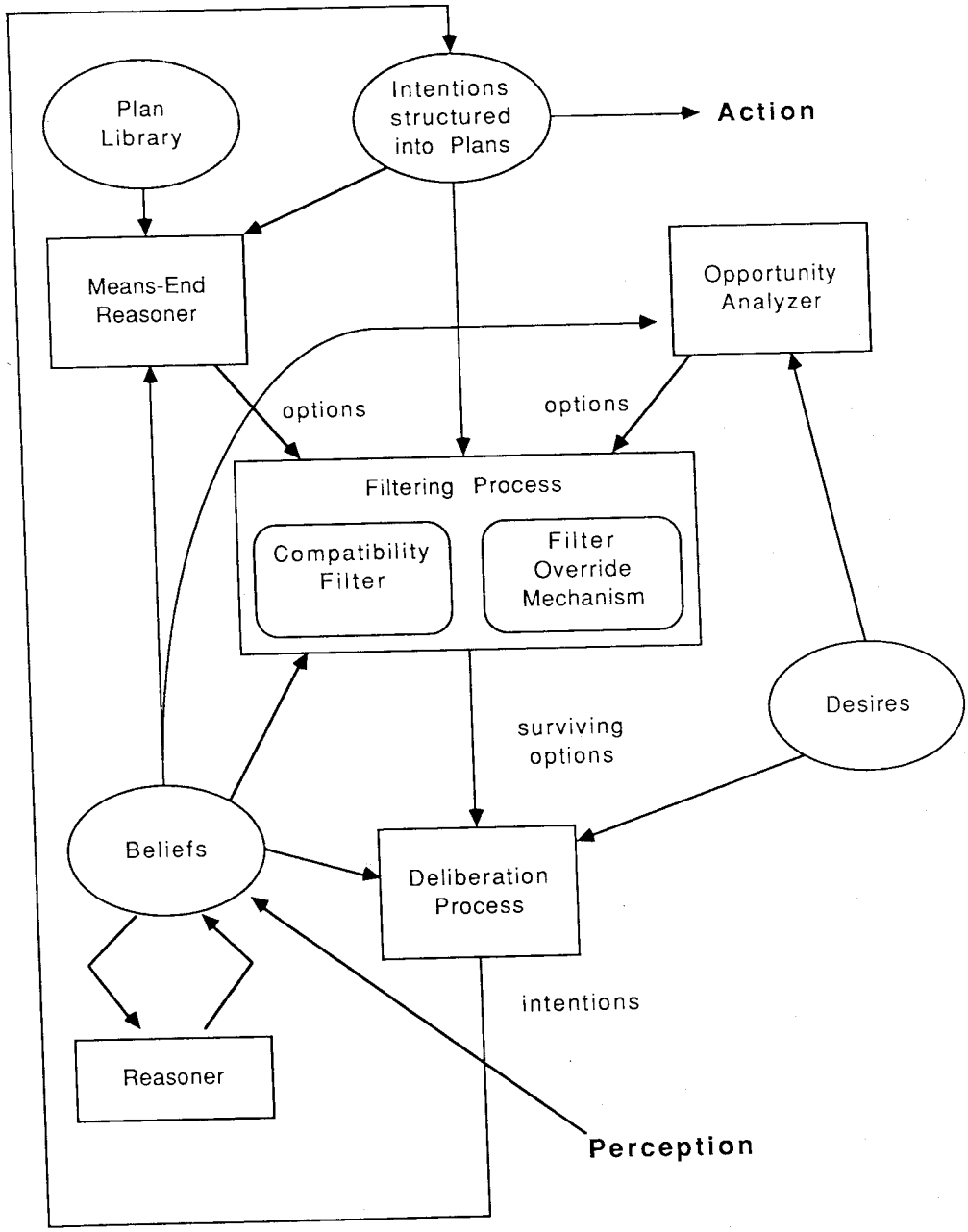

FIG. 1. An architecture for resource-bounded agents.

ticipated events. There thus exists a tension between the stability that plans must exhibit to play their role in focusing practical reasoning and the revocability that must also be inherent in them, given that they are formed on the basis of incomplete information about the future.

In the architecture of Fig. 1, this tension is mediated by the second component of the filtering process: the "filter override mechanism." The filter override mechanism encodes the agent's sensitivities to problems and opportunities in her environment — that is, the conditions under which some portion of her existing plans is to be suspended and weighed against some other option. The filter override mechanism operates in parallel with the compatibility filter. As we described above, an option that survives the compatibility filter is subject to consideration by the deliberation process: deliberation is not affected by the filter override mechanism. However, an option that does not survive the compatibility filter may still be subject to consideration if it triggers a filter override, i.e., if it satisfies one of the conditions encoded by override mechanism. If an option fails to survive the compatibility filter but does trigger a filter override, the intended act that is incompatible with the new option is *up for reconsideration*.

An agent's filter override mechanism must be carefully designed to embody the right degree of sensitivity to the problems and opportunities that arise in her environment. If the agent is overly sensitive, willing to reconsider her plans in response to every unanticipated event, then her plans will not serve sufficiently to limit the number of options about which she must deliberate. On the other hand, if the agent is not sensitive enough, she will fail to react to significant deviations from her expectations.

Consider what can happen to a proposed option. It may not survive the compatibility filter. Such an incompatible option may or may not trigger a filter override. If it does, the delibera-

TABLE 1. A taxonomy of practical-reasoning situations

Situation No.	Survives compatibility filter	Triggers filter override	Deliberation leads to change of plan	Deliberation *would have* led to change of plan
1	N	Y	Y	
2	N	Y	N	
3	N	N		N
4	N	N		Y
5	Y			

TABLE 2. Further complications

Situation No.	Survives compatibility filter	Triggers filter override	Deliberation leads to change of plan	Deliberation *would have* led to change of plan	Deliberation worthwhile
1*a*	N	Y	Y		Y
1*b*	N	Y	Y		N
4*a*	N	N		Y	Y
4*b*	N	N		Y	N

tion process will be invoked to decide between the incompatible option and the previously intended act that is now up for reconsideration. There is no guarantee that the agent will decide in favor of the new incompatible option; the result of deliberation may be a decision to maintain the previous intention.

If an incompatible option does not trigger a filter override, the agent does not deliberate about it. However, the designer (or other observer) of the agent may be able to determine what the result of properly functioning deliberation would have been, i.e., whether or not the agent's current beliefs and desires are such that deliberation would reasonably have led to her changing her plans to incorporate the new option.

Finally, it is of course possible that a proposed option will be compatible with the agent's existing plans, in which case it will be considered in deliberation. So we now have five possible situations, which are summarized in Table 1.

The next step is to note certain complications in situations 1 and 4. In both cases, these stem from the same fact: even deliberation that reasonably results in a change of intention takes time and so precludes other useful activities. This means that, in some instances of situation 1, the benefits achieved by the change in intention may be outweighed by the cost of deliberation. Similarly, in some instances of situation 4, the benefits that would have been obtained by the change in intention would have been outweighed by the deliberation required for this change. So there are two subcases of each of these situations, summarized in Table 2. In situation 1*a*, the change in intention has benefits that outweigh the cost of the extra deliberation, whereas in situation 1*b*, the cost of the deliberation outweighs the benefits of the change in intention. In situation 4*a*, the change that would have occurred would have had benefits outweighing the cost of the deliberation that would have been required, while in situation 4*b*, the opposite is the case.

5. Caution and boldness

Situations 1*b* and 2 have an important property in common: the agent engages in deliberation that is not on balance worth its cost. In contrast, in situation 4*a*, the agent fails to engage in deliberation that would have been worth its cost. Thus, an architecture that guaranteed that any agent embodying it would never be in these situations would be, at least in that respect, ideal.

Unfortunately, such an architecture is an impossibility.[8] Using the architecture we describe, one of the jobs of the robot designer is to construct the filter override mechanism so that, other things equal, it minimizes the frequency with which the agent will be in these situations.[9]

We can develop this last point and set the stage for some examples by introducing some terminology. When a proposed, but incompatible, option triggers a filter override, thereby leading to reconsideration of an existing intention, the agent is being *cautious*. When a proposed, but incompatible, option fails to trigger a filter override, the agent is being *bold*. What we have seen in our previous discussion is that sometimes caution pays and sometimes it doesn't; by the same token, sometimes boldness pays and sometimes it doesn't. In situation 1*a*, caution pays, whereas in situations 1*b* and 2, it doesn't. In situations 3 and 4*b*, boldness pays, whereas in situation 4*a*, it doesn't.

A filter override mechanism that results too often in cautious behavior that doesn't pay is *overly cautious*; one that results too often in bold behavior that doesn't pay is *overly bold*. In a well-designed agent, the filter override mechanism will be neither overly cautious nor overly bold.

Consider a robot Rosie, whose task is to repair computer

[8]Indeed, if such an architecture were possible, it would seem to have an odd consequence. Consider an agent who is disposed to be deliberate about an incompatible option when and only when that deliberation would lead to a worthwhile change. Such an agent might as well be designed to decide in favor of the new option whenever she is disposed to deliberate about it.

[9]In fact, one might ultimately want the filter override mechanism to be capable of being altered by the agent herself: if she realizes that she is spending too much time in fruitless deliberation, she should raise the sensitivity thresholds in the override mechanism; and, if she notices too many missed opportunities, she should lower the thresholds.

equipment. We present six different scenarios involving Rosie, scenarios that illustrate the situations described above in which an option fails to survive the compatibility filter. In each of the scenarios, we imagine that Rosie has been assigned several tasks, the first of which is to fix a malfunctioning video display on some terminal. We assume that Rosie does some means-end reasoning before setting off to do the repair: she determines that the best way to fix the problem is to replace the CRT, basing this decision upon a belief that CRTs burn out regularly as well as on an assumption that this is the cause of the malfunction. She thus brings a replacement tube along with her.

In the first scenario, Rosie arrives and discovers that the problem with the video display is caused by the contrast on the terminal being turned off. The opportunity analyzer proposes a new option: to fix the malfunction by simply turning up the contrast. This option is incompatible with Rosie's intention to fix the problem by replacing the CRT, yet she reconsiders her plan because her filter override mechanism has been triggered. Rosie's deliberation leads her to change her plan: she drops her intention to replace the CRT, and instead forms an intention to fix the malfunction by adjusting the contrast. Not only is this new option superior to the CRT-replacement, but it is sufficiently superior to outweigh the cost of reconsideration. After all, turning up the contrast is known by Rosie to be a significantly cheaper solution than replacing the CRT. So Rosie is being cautious and her caution pays (situation 1*a*).

In the second scenario, Rosie discovers that the existing CRT is repairable. As in the last example, Rosie is cautious and, in light of this new information, reconsiders her prior intention to replace the CRT. This involves weighing the pros and cons of replacement versus repair, a complicated exercise. Her deliberation results in a decision to repair rather than replace. And, indeed, repairing is a slightly better option. However, instead of deliberating, Rosie could have simply gone ahead with her intention to replace the CRT, and proceeded more quickly to her next task. Given this cost of her deliberation, her caution doesn't pay in this case (situation 1*b*).

Now suppose instead that replacing the CRT is the superior option, say, because the existing one is quite old and hence likely to breakdown again soon. Hence, when Rosie reconsiders, she decides not to change her prior intention, but instead to go ahead and replace the CRT. Here again Rosie is cautious, and her caution doesn't pay (situation 2).

In the remaining three scenarios, which illustrate Rosie's being bold, we suppose that, upon arrival, she discovers the presence of a spare CRT of a slightly different kind, one that she could use for the replacement, instead of the one she brought.

In the first case, despite this new information, Rosie does not reconsider her prior plan to replace the CRT. And, in fact, even had she reconsidered, she would have stuck with her prior plan, since the type of CRT she brought with her is superior. Rosie has been bold, and her boldness has paid off (situation 3).

Next, suppose that the opposite is true: had Rosie reconsidered, she would have found the new CRT to be slightly superior. Still, we can ask whether or not the deliberation that would have been required would have been worth it. In one case, the deliberation is relatively easy and does not interfere in any serious way with Rosie's other activities. In this case, then, Rosie's boldness doesn't pay (situation 4*a*). Alternatively, the deliberation would have precluded important activities of Rosie's, in which case, despite the slight superiority of the new CRT, Rosie's boldness pays (situation 4*b*).

This last pair of examples highlights an interesting fact about the difference between situations 4*a* and 4*b*. In both cases, the agent performs an action that is inferior to a known alternative: in both cases she would have favored the alternative, had she deliberated. So in both cases there is a kind of suboptimality. However, situation 4*b* differs from situation 4*a* in the following respect: in 4*b*, the combination of deliberation and the change of intention, taken together, is inferior to simply going ahead with the original intention. So it is no criticism of a well-designed agent that she ends up in situation 4*b*.

An agent instantiating a well-designed architecture, then, will tend to be in situations 1*a*, 3, and 4*b*, and to avoid situations 1*b*, 2, and 4*a*. So other things being equal, we want to design a filter override mechanism that has this effect. But, of course, there are limits to fine-tuning. We cannot expect even a well-designed architecture always to avoid situations 1*b*, 2, and 4*a*.

Consider Rosie. Suppose that CRTs are very expensive, and Rosie knows this. It might be a good strategy for her to reconsider an intention to replace a CRT when an alternative means is proposed. After all, such reconsideration will, on many occasions, save the cost of a new CRT. Of course, there may also be times when this strategy lands Rosie in situations of type 1*b* or 2, in which her caution doesn't pay: recall the case of the very old CRT. Nonetheless, this strategy might, on balance, be a good one. A more finely tuned filter, one that would be more successful in avoiding these undesirable situations, would run increased risks of ending up in situation 4*a*. As we try to avoid caution that doesn't pay, we run an increased risk of boldness that doesn't pay.

And, of course, the opposite is true as well: as we try to avoid boldness that doesn't pay, we run an increased risk of undesirable cautiousness. Suppose that the difference between one CRT and another is minimal. Then it might be a good strategy for Rosie not to reconsider her original intention to use a specific CRT replacement when an alternative replacement becomes available. This might be so even though such a strategy might sometimes land Rosie in situations of type 4*a*. If we try to fine-tune Rosie's strategy so as to avoid this risk, we shall simultaneously increase the risk of having her end up in situations 1*b* or 2.

6. Conclusion

We have presented the outlines of an architecture that can be used in the design of artificial agents who are, after all, resource-bounded. A key feature of this architecture is a filtering process that constrains the overall amount of practical reasoning necessary. The operation of this filtering process is based upon a theory of the functional roles of plans in practicall reasoning. While this is fairly abstract architecture, it does pose several specific design problems. In particular, procedures are needed for

- detecting threats to means-end coherence;
- proposing new options in light of perceived changes in the environment;
- monitoring prior plans in light of changes in belief;
- checking compatibility with prior plans; and
- overriding the compatibility filter.

Of course, there are other design problems, such as means-end analysis and the weighing of conflicting options, that are common to a wide range of architectures for rational agency.

Here we have highlighted problems specific to the architecture we are proposing.

Acknowledgments

This work has been partially supported by a gift from the Systems Development Foundation. For their discussion of this work, the authors are grateful to other members of the Rational Agency Project at the Center for the Study of Language and Information: Phil Cohen, Michael Georgeff, Kurt Konolige, Amy Lansky, and Ron Nash. The authors would also like to thank an anonymous reviewer.

ALBUS, J. S. 1981. Brains, behavior, and robotics. BYTE Books, Peterborough, NH.

BRATMAN, M. E. 1987. Intention, plans and practical reason, Harvard University Press, Cambridge, MA.

COHEN, P. R., and LEVESQUE, H. 1989. Persistence, intention, and commitment. *In* Intentions in communication. *Edited by* P. R. Cohen, J. Morgan, and M. E. Pollack. MIT Press, Cambridge, MA. In press.

DIFINETTI, B. 1975. Theory of probability. John Wiley and Sons, Inc., New York, NY.

DURFEE, E. H., and LESSER, V. R. 1986. Incremental planning to control a blackboard-based problem solver. Proceedings of the Fifth National Conference on Artificial Intelligence, Philadelphia, PA, pp. 58–64.

FIKES, R. E., and NILSSON, N. J. 1971. Strips: a new approach to the application of theorem proving to problem solving. Artificial Intelligence, **2**: 189–208.

GEORGEFF, M. P. 1987*a*. Actions, processes, and causality. *In* Reasoning about actions and plans: proceedings of the 1986 workshop. Morgan Kaufmann, Los Altos, CA, pp. 99–122.

———1987*b*. Planning. *In* Annual review of computer science, Vol. 2. *Edited by* J. Traub. Annual Reviews, Inc., Palo Alto, CA, pp. 359–400.

GEORGEFF, M. P., and LANSKY, A. L. 1987*a*. Reactive reasoning and planning. Proceedings of the Sixth National Conference on Artificial Intelligence, Seattle, WA, pp. 677–682.

———*Editors*. 1987*b*. Reasoning about actions and plans: proceedings of the 1986 Wwrkshop. Morgan Kaufmann, Los Altos, CA.

JEFFREY, R. 1983. The logic of decision. 2nd ed. University of Chicago Press, Chicago, IL.

KAELBLING, L. P. 1987. An architecture for intelligent reactive systems. *In* Reasoning about actions and plans: proceedings of the 1986 workshop. Morgan Kaufmann, Los Altos, CA, pp. 395–410.

KAUTZ, H. A., and VILAIN, M. 1986. Constraint propagation algorithms for temporal reasoning. Proceedings of the Fifth National Conference on Artificial Intelligence, Philadelphia, PA, pp. 377–382.

LANSKY, A. L. 1987. A representation of parallel activity based on events, structure, and causality. *In* Reasoning about actions and plans: proceedings of the 1986 workshop. Morgan Kaufmann, Los Altos, CA, pp. 123–159.

MCDERMOTT, D. 1985. Reasoning about plans. *In* Formal theories of the commonsense world. *Edited by* J. R. Hobbs and R. C. Moore. Ablex Publishing Company, Norwood, NJ, pp. 269–317.

SACERDOTI, E. D. 1977. A structure for plans and behavior. American Elsevier, New York, NY.

SAVAGE, L. J. 1972. The foundations of statistics. 2nd ed. Dover Press, New York, NY.

SCHMIDT, C. F. 1985. Partial provisional planning: some aspects of commonsense planning. *In* Formal theories of the commonsense world. *Edited by* J. R. Hobbs and R. C. Moore. Ablex Publishing Company, Norwood, NJ, pp. 227–250.

SIMON, H. 1957. Models of man. Macmillan Press, New York, NY.

SRIDHARAN, N. S., and BRESINA, J. 1982. Plan formation in large, realistic domains. Proceedings of the Fourth National Conference of the Canadian Society for Computational Studies of Intelligence, Saskatoon, Sask., pp. 12–18.

WILKINS, D. E. 1984. Domain-independent planning: representation and plan generation. Artificial Intelligence, **22**: 269–301.

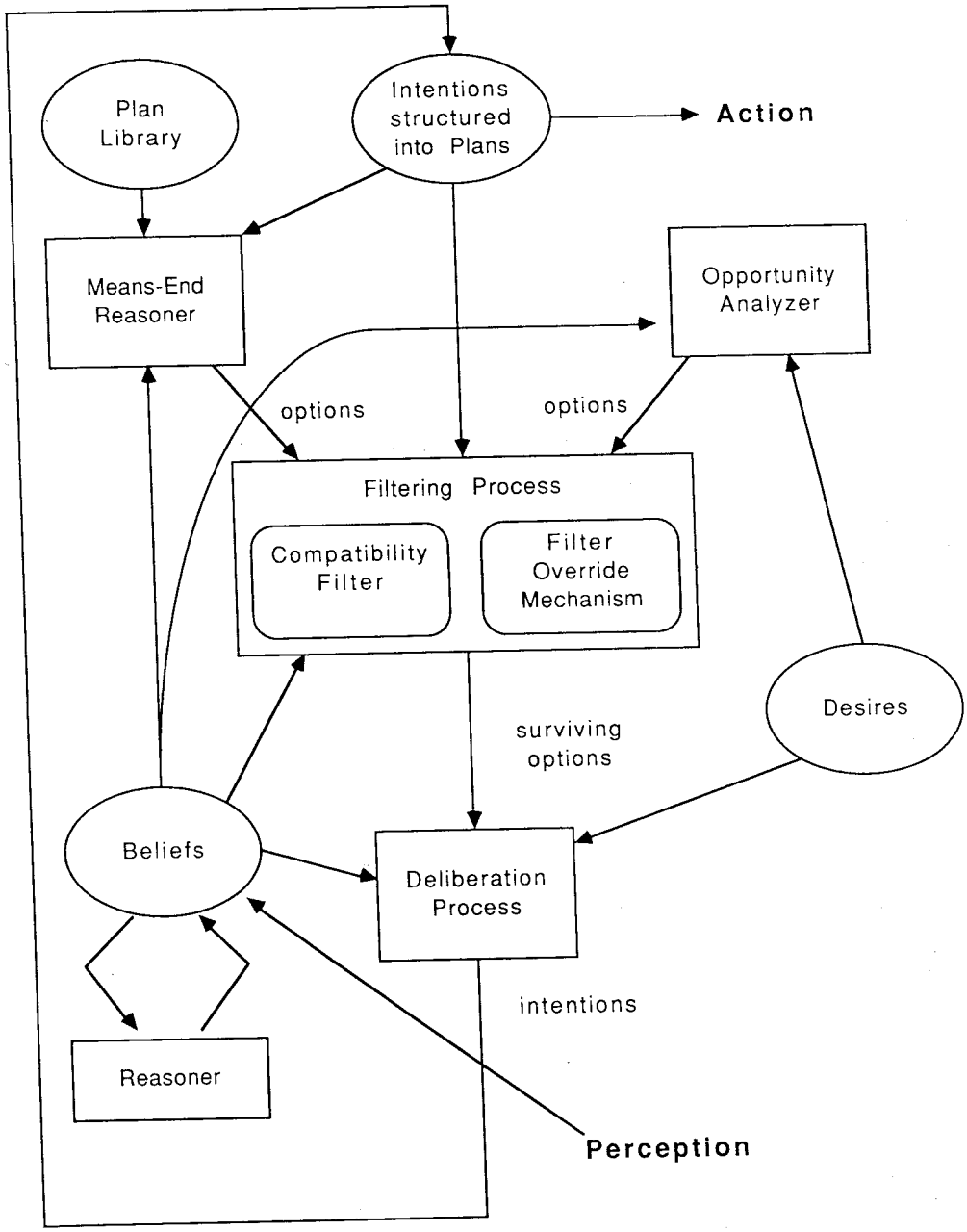

FIG. 1. An architecture for resource-bounded agents.

ticipated events. There thus exists a tension between the stability that plans must exhibit to play their role in focusing practical reasoning and the revocability that must also be inherent in them, given that they are formed on the basis of incomplete information about the future.

In the architecture of Fig. 1, this tension is mediated by the second component of the filtering process: the "filter override mechanism." The filter override mechanism encodes the agent's sensitivities to problems and opportunities in her environment — that is, the conditions under which some portion of her existing plans is to be suspended and weighed against some other option. The filter override mechanism operates in parallel with the compatibility filter. As we described above, an option that survives the compatibility filter is subject to consideration by the deliberation process: deliberation is not affected by the filter override mechanism. However, an option that does not survive the compatibility filter may still be subject to consideration if it triggers a filter override, i.e., if it satisfies one of the conditions encoded by override mechanism. If an option fails to survive the compatibility filter but does trigger a filter override, the intended act that is incompatible with the new option is *up for reconsideration*.

An agent's filter override mechanism must be carefully designed to embody the right degree of sensitivity to the problems and opportunities that arise in her environment. If the agent is overly sensitive, willing to reconsider her plans in response to every unanticipated event, then her plans will not serve sufficiently to limit the number of options about which she must deliberate. On the other hand, if the agent is not sensitive enough, she will fail to react to significant deviations from her expectations.

Consider what can happen to a proposed option. It may not survive the compatibility filter. Such an incompatible option may or may not trigger a filter override. If it does, the delibera-

TABLE 1. A taxonomy of practical-reasoning situations

Situation No.	Survives compatibility filter	Triggers filter override	Deliberation leads to change of plan	Deliberation *would have* led to change of plan
1	N	Y	Y	
2	N	Y	N	
3	N	N		N
4	N	N		Y
5	Y			

TABLE 2. Further complications

Situation No.	Survives compatibility filter	Triggers filter override	Deliberation leads to change of plan	Deliberation *would have* led to change of plan	Deliberation worthwhile
1*a*	N	Y	Y		Y
1*b*	N	Y	Y		N
4*a*	N	N		Y	Y
4*b*	N	N		Y	N

tion process will be invoked to decide between the incompatible option and the previously intended act that is now up for reconsideration. There is no guarantee that the agent will decide in favor of the new incompatible option; the result of deliberation may be a decision to maintain the previous intention.

If an incompatible option does not trigger a filter override, the agent does not deliberate about it. However, the designer (or other observer) of the agent may be able to determine what the result of properly functioning deliberation would have been, i.e., whether or not the agent's current beliefs and desires are such that deliberation would reasonably have led to her changing her plans to incorporate the new option.

Finally, it is of course possible that a proposed option will be compatible with the agent's existing plans, in which case it will be considered in deliberation. So we now have five possible situations, which are summarized in Table 1.

The next step is to note certain complications in situations 1 and 4. In both cases, these stem from the same fact: even deliberation that reasonably results in a change of intention takes time and so precludes other useful activities. This means that, in some instances of situation 1, the benefits achieved by the change in intention may be outweighed by the cost of deliberation. Similarly, in some instances of situation 4, the benefits that would have been obtained by the change in intention would have been outweighed by the deliberation required for this change. So there are two subcases of each of these situations, summarized in Table 2. In situation 1*a*, the change in intention has benefits that outweigh the cost of the extra deliberation, whereas in situation 1*b*, the cost of the deliberation outweighs the benefits of the change in intention. In situation 4*a*, the change that would have occurred would have had benefits outweighing the cost of the deliberation that would have been required, while in situation 4*b*, the opposite is the case.

5. Caution and boldness

Situations 1*b* and 2 have an important property in common: the agent engages in deliberation that is not on balance worth its cost. In contrast, in situation 4*a*, the agent fails to engage in deliberation that would have been worth its cost. Thus, an architecture that guaranteed that any agent embodying it would never be in these situations would be, at least in that respect, ideal.

Unfortunately, such an architecture is an impossibility.[8] Using the architecture we describe, one of the jobs of the robot designer is to construct the filter override mechanism so that, other things equal, it minimizes the frequency with which the agent will be in these situations.[9]

We can develop this last point and set the stage for some examples by introducing some terminology. When a proposed, but incompatible, option triggers a filter override, thereby leading to reconsideration of an existing intention, the agent is being *cautious*. When a proposed, but incompatible, option fails to trigger a filter override, the agent is being *bold*. What we have seen in our previous discussion is that sometimes caution pays and sometimes it doesn't; by the same token, sometimes boldness pays and sometimes it doesn't. In situation 1*a*, caution pays, whereas in situations 1*b* and 2, it doesn't. In situations 3 and 4*b*, boldness pays, whereas in situation 4*a*, it doesn't.

A filter override mechanism that results too often in cautious behavior that doesn't pay is *overly cautious*; one that results too often in bold behavior that doesn't pay is *overly bold*. In a well-designed agent, the filter override mechanism will be neither overly cautious nor overly bold.

Consider a robot Rosie, whose task is to repair computer

[8]Indeed, if such an architecture were possible, it would seem to have an odd consequence. Consider an agent who is disposed to be deliberate about an incompatible option when and only when that deliberation would lead to a worthwhile change. Such an agent might as well be designed to decide in favor of the new option whenever she is disposed to deliberate about it.

[9]In fact, one might ultimately want the filter override mechanism to be capable of being altered by the agent herself: if she realizes that she is spending too much time in fruitless deliberation, she should raise the sensitivity thresholds in the override mechanism; and, if she notices too many missed opportunities, she should lower the thresholds.