Towards a formalization of the formation and practicalization of goals

Damian Wassel
University of Michigan

Richmond Thomason
University of Michigan

Abstract

The literature on planning agents (and, in particular, BDI agents) has concentrated on means-end reasoning – that is, on determining a sequence of actions to achieve a given goal. Relatively little attention has been given to the problem of practicalizing desires – that is, arriving at reasonable goals from raw desires. When processes of goal formation are taken into account, this is usually done procedurally, as part of the architecture of the planning agent.

Work on planning systems has been enabled and transformed by logical formalization of means-end reasoning. We believe that goal formation can benefit from a logical treatment, even though reasoning about desires may be somewhat less tractable than causal reasoning. As a first step at providing a formalization, we adopt Horty's treatment of reasons (reasons for action as well as reasons for belief) in terms of prioritized defaults. This, we argue, provides a basis for formalizing motivations, and for converting motivations and recognized opportunities into provisional goals.

Introduction

Most work on on "rational agents"—agents that reason in a principled way about what to do—has concentrated on the formation of new intentions, based on beliefs, desires, and old intentions. Because planning and implementations of planning systems represent one of the most successful integrations of logical theory with AI applications—and in view of the fact that in many applications, appropriate desires or goals for planning can be regarded as inputs to the planning agent—this is not surprising.

But reasoning about what to do is not limited to planning. We also reason about what goals to have, and a fully autonomous agent would need the capability to generate appropriate goals. The relatively little work that has been done on this aspect of rational agency is usefully surveyed in (Hawes 2011).

Hawes distinguishes goal generation from *goal management,* which he says selects which of the generated goals "are subsequently allowed to influence the systems behaviour," and uses "motive management framework" for a combined system that performs both tasks. We prefer to organize things differently. We separate planning and intention formation in advance of execution from the maintenance and adjustment of existing plans and from their scheduling, online monitoring, and execution. (Nilsson 1994) provides a treatment of execution with his teleoreactive approach; for plan monitoring, see, for instance, (Micalizio and Torasso 2007).

We view goal generation as the selection of provisional goals. This process then automatically triggers a planning process, with the presumption that planning will lead to an intention. We think of intentions in the usual way, as a commitment to a course of action (See (Bratman, Israel, and Pollack 1988), (Cohen and Levesque 1986).) But the outcome of a deliberative process involving planning is only presumptive. Once plans have been formed, they need to be evaluated. A plan may not lead to an intention because there are better competing plans, because of its inherent costs, or because it competes with other plans that satisfy other desires. But in simple cases the path from a practical goal to an intention is automatic.

Practicalization is the entire reasoning process that leads from desires and beliefs to intentions. We divide practicalization into three steps. (1) Desires, together with perceived opportunities, give rise to *provisional goals*. (2) Means-end reasoning or planning produces alternative ways of achieving these goals. (3) A process of plan evaluation selects the plans that are worthy of conversion to intentions. In this paper, we concentrate on Step (1), which we feel is the neglected part of this process. The literature on Step (2) is enormous, and the literature on Step (3) is substantial. The filtering process described in (Pollack 1992), for instance, is devoted to the evaluation of new options, mainly in light of plans to which the agent is already committed.

This paper proposes how to formalize this reasoning in a general format that includes not only conflicts with existing plans and measurable costs of new plans, but more subtle reconciliations of competing desires with the realistic constraints of planning.

Previous work

(Georgeff and Ingrand 1989) describe a procedural reasoning system for an agent, inspired by BDI architectures, that is capable of forming and executing plans. In a section on "The establishment and dropping of goals", they discuss only the establishment of intentions; apparently their system, like any planning system, will produce subgoals (which they call "operational goals"), but does not create "intrinsic goals."

Like any procedures, the procedures that comprise know-how in this system include: (1) an action (which may be arbitrarily complex), and (2) "an invocation condition" that serves to trigger the action. Invocation conditions can include goals as well as declarative conditions on situations. Although there is a mechanism for the activation of these procedures, apparently the entire set of procedures has to be hand-coded into this system, as with expert systems, although there is no reason in principle why these procedures could not be learned. Later in this paper, we will show how logical techniques can be used to derive invocation conditions, which we call *opportunities*.

(Coddington et al. 2005) is unusual in providing explicit representations of certain desires, in the form of *drives*. In later work, (Coddington 2006; 2007), Alexandra Coddington deals with what she calls *motivations*. For her, motivations have two parts: a value (a fluent representing the strength of motivation) and an importance, representing the centrality of of the motivation. Associated goals are activated when a motivation's value exceeds a certain threshold. This process of goal generation is then integrated with an autonomous planning system.

The work we have surveyed provides for some simple models of desires and of processes of goal generation, and shows that these processes can be integrated into BDI agent architectures. We know of no work, though, that attempts to use logic to specify the associated reasoning, as, for instance, (Wooldridge 2000) does with the core reasoning of BDI agents. Providing such a logical account of goal generation is what we attempt to do in this paper. No doubt, reasoning about desires presents more of a challenge to logical formalization than many other kinds of reasoning. What we will do here is no more than a beginning.

The formalization of desires

In (Horty 2012), John F. Horty argues persuasively that reasons can be formalized as (normal) prioritized defaults. This formalization captures many important features of reasons: (1) they are defeasible, but (2) in the absence of contrary reasons, they yield conclusions, (3) they can conflict, and conflicts need to be applied in reasoning with them, (4) sometimes in resolving conflicts between reasons, one reason is clearly better or stronger than another.

This formalization is explicitly meant to apply not only to epistemic reasons (reasons for accepting something as true) but to practical reasons (reasons for doing something).

Although not every reason for doing something is a motive (an agent may not be motivated by a reason—even a very good reason—for an action), many are. Also, reasoning with motives exhibits the four general properties that led Horty to explain them as prioritized defaults. We can illustrate these with the following simple examples.

1. Hunger is a motive, and a powerful one. But there are many reasons why a hungry person may not eat. For a while at least, hunger may be overridden by a desire to lose weight.
2. In the absence of any motive to the contrary, and in the presence of food, we'd expect a hungry person to eat.
3. Ambition can conflict with laziness.
4. Promise-keeping is in general a better motive than laziness. It may not always be stronger.

We will need a language for formalizing motives that is also able to formalize planning. For this purpose, we will use a version of the Situation Calculus as the monotonic basis for a prioritized default theory. But here, defaults are only used to formalize desires, and so cannot, for instance, be used to represent epistemic defaults. In particular, these defaults can't be used to solve the frame problem, and in fact, to keep things simple, it would be better to imagine a monotonic solution to the frame problem.[1]

A motive, in our language, will be a normal default rule $\delta = \phi \rightsquigarrow \psi$, where ϕ and ψ are formulas in the first-order language of the Situation Calculus. We call the left-hand side of such a rule its *premise* and the right-hand side of such a rule its *conclusion*. In keeping with Horty, we take a default rule to express a favoring relationship between premise and conclusion. In particular, since our interests are in the practical domain, we take a default rule to mean that the premise makes the conclusion desirable. For example, let i be some agent, and let $\mathbf{H}$ and $\mathbf{E}$ stand-in for formulas of the Situation Calculus that express, respectively, that i is now hungry, and that i eats in the near future. Then the default, $\mathbf{H} \rightsquigarrow \mathbf{E}$, means that i's present hunger makes it desirable that i eat in the near future. Throughout, we use defaults of the form $\rightsquigarrow \psi$ as shorthand for defaults of the form $\top \rightsquigarrow \psi$. The set of defaults of a prioritized default theory are ordered by a strict partial ordering $\prec$. We will need to have general motives, but defaults are rules, not formulas, and cannot themselves be universally quantified. This problem is solved by using rule schemata, where a schema denotes all its substitution instances.

A default theory DT of motives is no different from an ordinary prioritized default theory: it consists of a set of first-order formulas (the monotonic axioms), a set of default rules, and a partial ordering over the defaults. As usual with default theories, DT will in general not yield a unique set of consequences, because some conflicts between motives may be unresolved. The standard definition of logical consequence is replaced by a definition of the set of *extensions* that is associated with DT; each extension represents a reasonable set of conclusions from DT, from which no further conclusions may be drawn.[2] We can think of choices between these extensions as alternatives that are available to the agent, but not as reasoned conclusions; perhaps the choice is made arbitrarily. Imagine someone in bed, moved equally by ambition, which tells him to get up, and laziness,

[1] Epistemic defaults and practical defaults play quite different roles in practical reasoning, and if both are present, they need to be kept separate, to avoid fallacies of "wishful thinking." See (Thomason 2000).

[2] In this paper we have no need for a precise definition of an extension. We direct interested readers to (Horty 2012) for such a definition.

which tells him to stay in bed. He will have to resolve this conflict somehow, but can't do that in his current motivational state.

This formalization of motives brings with it a certain amount of complexity—multiple extensions, and a definition of extensions that is by no means simple. On the other hand, it has many expressive advantages over the very simple formalizations of motivation using threshold values of certain fluents. These would be desires of the form $\rightsquigarrow f(s) \geqslant c$, where s is a variable ranging over situations. But there are conditional desires, desires that wouldn't naturally be formalized in terms of threshold values, and desires directed at states that are logically complex. Consider the following examples.

1. If I'm going to miss my class, someone else needs to teach it.
2. If the airports in Chicago are closed, I don't want to be in Chicago.
3. If there are no flights to Chicago, I want to go to Chicago by bus or by train.
4. If I'll have to stay late at work, my wife needs to know that I'll be home late.

These, of course, are highly specific, occasional desires, rather than generic, standing desires like drives. But, if we are to formalize and reason with desires, and if generic desires provide reasons for specific desires, we need formalizations of both.

This more expressive language for formalizing desires provides a solution to another problematic aspect of the threshold value treatment of drives. Suppose that a robot's need to energize its batteries is formalized as a production that proposes the goal of recharging when battery capacity falls below 50%. This simple formalization couldn't provide a reason for the robot to recharge if it had nothing else important to do, was near a recharging station, and had a battery capacity of, say 90%. A default

$$(1) \quad \textsc{Capacity}(s) < 100 \rightsquigarrow \exists s'[\textsc{Near-Fut}(s', s) \wedge \textsc{Capacity}(s') > \textsc{Capacity}(s)]$$

with a priority that becomes larger as the value of $\textsc{Capacity}(s)$ decreases, will account for a much larger variety of rational recharging behaviors. Here $\textsc{Near-Fut}(s', s)$ is a relation over situations that holds just in case s' is in the near future of s, where 'near future' is given some context-sensitive interpretation. We take this default to mean that when the batteries are below maximum capacity, it is desirable to increase their capacity.

Using prioritized defaults to formalize desires puts a certain amount of pressure on the priorities and their use in constructing extensions. The prioritization of desires seems to involve many factors, such as emotional strength, a sense of the values of various outcomes, and appreciation of risk. It may well be an oversimplification to assume that these factors are combined seamlessly in a single ordering relation.[3]

[3]Horty's idea of a *variable priority default theory*, which provides an explicit way of representing and reasoning about priorities, may help with this problem.

Also, Reiter's extension construction doesn't seem to fit some plausible cases in which competing desires are reconciled. Extensions are formed by looking at all the results of winner-take-all competitions between competing defaults, but the reconciliation of desires may involve compromise. Consider, for example, an overweight but dieting diner, who is tempted by a slice of cheesecake. We can imagine the diner motivated in equal measure by a drive to eat the cheesecake, and a drive to maintain the diet. In this example, a reasonable thing for the diner to do might be to compromise, and opt for some low-calorie dessert, even though the diner doesn't directly desire this outcome. We see this sort of compromise as reasonable, because it doesn't altogether wreck the diet, and it helps stave off cravings for the cheesecake. However, we're not readily able to formalize this sort of reasoning using normal defaults.[4]

Here, we simply note that such examples may require fundamental changes in the logical treatment of desires as defaults, and leave it as an open problem.

Our provisional account of motives formalizes them as prioritized defaults. A motivated agent and the relevant facts about its environment, then, would be formalized as a prioritized default theory, an ordered triple of the form $\langle \mathcal{W}, \mathcal{D}, \prec \rangle$, where $\mathcal{W}$ is a set of monotonic axioms, $\mathcal{D}$ is a set of defaults, and $\prec$ is a strict partial ordering over $\mathcal{D}$. With this basis, we now turn to the problem of goal formation.

Motive and opportunity

Detective fiction teaches us the importance of motive and opportunity. Typically a provisional goal—something that leads us to ask the question "How could I go about doing this"—will emerge from the combination of a moving desire, a *motive* which can either be generic or occasional, with a recognized opportunity for furthering the motive. Certain very important human choices may not arise in this way. Consider choices of long-term commitment like selecting one's career or life partner. However, we think most mundane goals do come about in this manner.

The following simple example illustrate this process.

Example 1. Unpacking my bag in a hotel, I find that I forgot to pack toothpaste. Later, walking to an ap-

[4]Let $\mathbf{C}$ stand in for a formula of the Situation Calculus that expresses that the diner eats the cheesecake in the near future, and $\mathbf{D}$ stand in for a formula of the Situation Calculus saying that the diner keeps to the diet in the near future. Then we can represent the drive to eat the cheesecake with $\delta_1 = \rightsquigarrow \mathbf{C}$ and the drive to keep to the diet with $\delta_2 = \rightsquigarrow \mathbf{D}$. Then we can represent a simplified version of this case with the following default theory: $DT_1 = \langle \emptyset, \{\delta_1, \delta_2\}, \emptyset \rangle$. Since the two defaults have equal priority, DT_1 has two extensions, one in which δ_1 is the operative motive, $E_1 = \{\mathbf{C}\}$, and another, $E_2 = \{\mathbf{D}\}$, in which δ_2 is the operative motive. The compromise of eating the low-calorie desert is nowhere to be found among these extensions. To generate an extension in which this did appear, some formula of the Situation Calculus saying that the diner eats the low-calorie desert would have to appear as (part of) the conclusion of some default in the theory. This, of course, is problematic, because by supposition the diner has no drive to eat the low-calorie desert. This highlights a limitation of formalizing motivational sets with default logic.

pointment, I pass a drugstore. On the spot, I form an intention to buy toothpaste, and act on it.

This examples raise two issues: (i) How do standing desires give rise to an operative desire to have toothpaste? (ii) How does the agent recognize that the current situation affords an opportunity? We consider each of these in turn.

Producing an operative desire from standing desires.

If I had no standing desire to brush my teeth regularly, there would be no motive in Example 1 to have toothpaste. This standing desire, of course, is not a drive, but it's a habit that I learned at a young age. Let's suppose, then, that because of childhood training and habit, I have a generic desire to brush my teeth in the morning and at bedtime. The second desire might be formalized by the following default schema.

$$(2) \quad [\textsc{Bedtime}(s) \wedge \neg\exists s'[\textsc{Evening}(s') \wedge s' < s \wedge \textsc{Same-Day}(s, s') \wedge \textbf{Do}(\textsc{Brush-Teeth}, s')] \rightsquigarrow \exists s''[\textsc{Near-Fut}(s'', s) \wedge \textbf{Do}(\textsc{Brush-Teeth}, s'')]$$

This default schema represents a standing desire, which we can assume is activated every evening at bedtime with moderate priority. In this example, the particular instance of this desire-schema plays the role of motive.

Let's suppose that my general knowledge about actions—the same knowledge that is used for planning—includes preconditions for the action BRUSH-TEETH, and that among these preconditions are having toothpaste, having a toothbrush, and being colocated with these items. Let's further assume that the formalized knowledge conditions for actions distinguish *resource* preconditions from other preconditions: among other things, resource preconditions generally involve material requirements like tools, money, and fuel. Then from a standing desire like (2) we can recover derived, secondary desires for the resources that are required by standing goals.

These secondary desires can be derived by metarules, which permit defaults to be inferred from other defaults. The resource requirement metarule would look like this.

If $\phi(s) \rightsquigarrow \exists s'[\textsc{Near-Fut}(s', s) \wedge \textbf{Do}(a, s')]$
is a desire and a requires resource r, then
$\phi(s) \rightsquigarrow \exists s'[\textsc{Near-Fut}(s', s) \wedge \textbf{Have}(r, s')]$
is also a desire, with the same priority.

Of course, desires of this kind are seldom activated—I go through most of my days without a thought about my need for toothpaste in the evening. We need some connection between desires for resources and experiences, so that certain observations will register as *lacks*. This connection, we suppose, is a matter of memory and attention, and as such may not be appropriate for logical formalization.

In our example, noticing that I forgot the toothpaste, together with my standing desire to brush my teeth in the evening, the fact that toothpaste is a required resource for this action, and a certain amount of attention to the connection between the missing toothpaste and a standing goal, generates a provisional goal of acquiring toothpaste. Let's suppose that this plan fails to produce an intention, because I don't know where to get toothpaste and don't have time to find out. In the next section, we'll discuss how these things can produce an outlook for certain opportunities.

A logical treatment can also be of some help in explaining why in this example I recognize an opportunity when I happen to pass a drugstore. The idea is that desires lead to provisional goals, provisional goals invoke a planning process, lacks can be recovered from a failed plan, and the observation of something that satisfies a lack amounts to the recognition of an opportunity.

To formalize this idea, we will generalize plan verification from an attempt to provide a proof that performing an action will achieve a goal to an attempt to identify the resources that would be needed to provide a proof of the goal. We imagine a process that, like ordinary plan verification, uses the preconditions of actions to create subgoals, but that makes rough estimates of the feasibility of achieving each precondition. If the feasibility of a resource precondition is high, the plan verification process proceeds as usual to look for actions that will achieve the precondition. If the feasibility is low, the process stops and simply lists the precondition as an obstacle to achieving the goal.

Epistemological abduction—reasoning from observations to a plausible explanation—can be regarded as the identification of additional likely assumptions that would allow the observation to be proved from available knowledge.[5] Here, we are thinking of *practical* abduction as the identification of likely subgoals that would need to be satisfied to achieve a frustrated goal. In Example 1, my plan to obtain toothpaste might get as far as considering the likely actions of buying toothpaste in a grocery store, in a convenience store, and in a drugstore. This would then produce the subgoal of knowing where a grocery store is or knowing where a convenience store is or knowing where a drugstore is. (To simplify the formalization, let's consider the simpler subgoal of knowing where a drugstore is.) At this point, when I find none of the knowledge conditions is satisfied and I realize I don't have time to address this problem, a lack has been identified. This leads to the creation of a new desire to address the lack, which can be formalized as follows.

$$(4) \quad \rightsquigarrow \exists s[\textsc{Near-Fut}(s, \textsc{Now})] \wedge \exists x \exists y[\textsc{Drugstore}(x) \wedge \textsc{Know}(^\wedge\textsc{Located}(x, y), s)]$$

Here '$\textsc{Now}$' is a constant which refers to the present situation.

In this way, a secondary desire is created by logical reasoning, using the same knowledge resources that are needed for ordinary planning, from a failed attempt to form a plan to satisfy a short-term primary desire. We assume that the secondary desire is linked to the primary desire that induced it. This linkage is characterized by certain relationships between the desires, among them that the primary desire cannot be fulfilled (at least in the course of carrying out the current operative plan) without fulfilling the secondary desire; any desire that takes priority over the primary desire takes priority over the secondary desire; and if the primary desire is somehow nullified, the secondary desire is nullified. Most

[5] For a discussion of this picture of abduction, and a description of an efficient algorithm that could be adapted to our present purpose, see (Stickel 1991).

importantly, if the secondary desire is fulfilled, the obstacle to fulfilling the primary desire is removed, and the agent may proceed with the previously frustrated operative plan.

Any autonomous planning agent needs to have a means of recognizing whether goals have been achieved; this mechanism is needed in order to tell when an intention can be dropped. The same mechanism can be used to monitor whether activated desires have been satisfied. The recognition that (4) has been satisfied, together with a link to the primary desire (2), reactivates the frustrated plan, which can then be completed, converted to an intention, and acted upon.

Competing desires and planning architectures

> *Example 2.* An American computer scientist, M, is invited to give a talk at an IJCAI meeting. She forms the provisional goal of attending the meeting and giving the talk. The goal is provisional, because of potential conflicts with other desires, some of them reflecting previous commitments. Eventually she produces a plan that seems acceptable in view of all the relevant desires, and adopts it as an intention.[6]

Let's suppose that M's deliberation begins with an invitation to address a plenary session of IJCAI. Before this, because of a family reunion and a move to a new location and job at about the same time as the meeting, M had dismissed the idea of attending IJCAI. We can suppose that there is a standing desire to attend meetings involving an invitation that promotes professional advancement, and that M believes that this is such an invitation. This leads to a desire, in the form of a default, to attend the meeting. But the desire is only provisional, because it has to be reconciled with the preexisting desire (actually, an intention) to attend the family reunion (which, we suppose, is at the same time as the IJCAI meeting). Furthermore, if M attends the meeting, she will miss part of the move, making it harder on her family, and will start her new job jetlagged.

Pollack used this example to illustrate the IRMA planning architecture (Bratman, Israel, and Pollack 1988; Pollack 1992), based on the idea that intended plans act as a filter on the adoption of new goals, and on plans based on these goals. This example is far too complex to allow the level of formal detail we attempted in Example 1. But we discuss it here to indicate in broad terms how our formalization of desires can provide a logical model of the IRMA filtration process.

Typically desires do not conflict directly, but an indirect conflict will emerge in the process of provisional planning. So to identify incompatible desires we must compare plans based on these desires. Using ideas in (Thomason 2000), we can combine the formation of extensions from a default theory with planning. A *plan extension* is a default extension that contains a plan to achieve the desires contained in this extension. Two desires are *practically incompatible* if no plan extension satisfies them both. In our example, we can assume that the desire to attend the reunion and the desire to attend IJCAI are incompatible in this sense, because of the time conflict. Therefore, no intention can emerge from deliberation that satisfies both desires; which desires are chosen and which are discarded will be a matter of priority.

We can formalize the idea of (Bratman, Israel, and Pollack 1988) that commitments to plans act as a filter on future plans and goals by boosting the priority of adopted goals in comparison with goals that are only provisional; but other factors can lead to a provisional goal having overall higher priority.[7]

In this example, although the priority of the desire to attend the reunion is enhanced because of a preexisting intention, we suppose that the motivation for professional advancement trumps it.

The resolution of the various desires having to do with the new job and the move are more complicated, but we can suppose that some of these (such as the very high priority desire to move household goods) can be achieved in a single plan that includes attending IJCAI, and that the remaining desires (such as helping with the move) have relatively low priority in comparison with professional advancement. This makes it plausible, at least, that a plan extension incorporating the move and the meeting would have relatively high priority.

Subgoaling

The method of subgoaling is available to any planning agent as a way of forming goals: as the agent forms a plan to achieve a goal situation ϕ_0, a partial plan is formed, containing an action, which, if performed, will achieve the goal. To be performed, the joint preconditions of this action must hold in the situation in which the action is performed. A provisional goal of achieving a situation satisfying these conditions is formed, and the planning process is repeated. If the agent eventually commits to the resulting plan, the concomitant intentions will include not only the ultimate goal, but the subgoals that are included in the plan. As they say, "to will the end is to will the means." Why can't the method of subgoaling account for the new goals that are formed in Examples 1 and 2? If these new goals were subgoals, the apparatus we have introduced would be redundant and unnecessary.

In Example 1, subgoaling can't account for the intention to buy a toothbrush in a drugstore because on the BDI model of plan formation a plan must be completed for an intention to be adopted. Assuming that knowing the location of a nearby drugstore is a precondition of buying a toothbrush in a store, the plan in this example will fail because the precondition fails and the agent has not completed a plan to remedy the knowledge deficiency. At this point, according to the theory we propose, a desire to buy a toothbrush in a drugstore is formed, and it is this desire that makes the recognition of an opportunity possible. This is very like the formation of subgoals, but differs in not requiring the formation of a com-

[6]This example is based on Martha Pollack's trip to IJCAI, described in (Pollack 1992). But we have felt free to imagine some details to enhance the example for our purposes.

[7]To implement this idea, we would need a mechanism for maintaining and updating priorities among desires. We do not expect this to be easy, but hope that existing work on reasoning about preferences can be used here.

plete plan. And this mechanism requires a way to formalize the relevant desire.

In Example 2, subgoaling obviously can't produce the goal of attending IJCAI. Instead, the goal emerges from a complex deliberation process involving reconciliation of competing desires and the practicalities of planning. Some desires are discarded as goals in this process; others may be satisfied or partially accommodated. Although we aren't able to provide formal details here, we have tried to make it plausible that prioritized defaults provide a promising framework for formalizing this sort of reasoning.

Conclusion

We have shown that a formalization of desires, along the lines suggested by Horty, can provide a logical account of how secondary desires might be derived from primary desires and converted to provisional goals by a planning agent. This, of course, raises many problems. Other than the ones we have mentioned, the largest is whether this formalization could be usefully applied, not just to toy domains, but to complex and realistic domains of the sort that BDI agents have been adapted to, such as autonomous spacecraft.

We think we can claim, however, that what we have done here suggests a way of extending the logical models of BDI agents into the new and underexplored territory of motives and goal formation.

References

Bratman, M. E.; Israel, D.; and Pollack, M. 1988. Plans and resource-bounded practical reasoning. *Computational Intelligence* 4:349–355.

Coddington, A.; Fox, M.; Gough, J.; Long, D.; and Serina, I. 2005. MADbot: A motivated and goal directed robot. In Manuela M. Veloso, S. K., ed., *The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference*, 1680–1681. Menlo Park, California: AAAI Press.

Coddington, A. M. 2006. Motivations for madbot: a motivated and goal directed robot. *Proceedings of the 25th Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG 2006)* 39–46.

Coddington, A. M. 2007. Integrating motivations with planning. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '07, 850–852. New York, NY, USA: ACM.

Cohen, P. R., and Levesque, H. J. 1986. Persistence, intention, and commitment. In Georgeff, M. P., and Lansky, A., eds., *Reasoning about Actions and Plans*. Los Altos, California: Morgan Kaufmann. 297–340.

Georgeff, M. P., and Ingrand, F. F. 1989. Decision-making in an embedded reasoning system. In Sridharan, N., ed., *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 972–978. San Mateo, CA: Morgan Kaufmann.

Hawes, N. 2011. A survey of motivational frameworks for intelligent systems. *Artificial Intelligence* 175(5):1020–1036.

Horty, J. F. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.

Micalizio, R., and Torasso, P. 2007. Plan diagnosis and agent diagnosis in multi-agent systems. In Basili, R., ed., *Artificial Intelligence and Human-Oriented Computing: 10th Congress of the Italian Association for Artificial Intelligence*. Berlin: Springer Verlag. 434–446. LNCS, vol. 4733.

Nilsson, N. J. 1994. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research* 1:139–158.

Pollack, M. 1992. The uses of plans. *Artificial Intelligence* 57(1):43–68.

Stickel, M. E. 1991. A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. *Annals of Mathematical Artificial Intelligence* 4:89–105.

Thomason, R. H. 2000. Desires and defaults: A framework for planning with inferred goals. In Cohn, A. G.; Giunchiglia, F.; and Selman, B., eds., *KR2000: Principles of Knowledge Representation and Reasoning*, 702–713. San Francisco: Morgan Kaufmann.

Wooldridge, M. J. 2000. *Reasoning about Rational Agents*. Cambridge, England: Cambridge University Press.