Logical Theories of Intention and the Database Perspective

Yoav Shoham

Abstract While logical theories of information attitudes, such as knowledge, certainty and belief, have flourished in the past two decades, formalization of other facets of rational behavior have lagged behind significantly. One intriguing line of research concerns the concept of intention. I will discuss one approach to tackling the notion within a logical framework, based on a database perspective.

1 Introduction

Logics of rational agency attempt to capture, in logic, various facets of human mental state, and posit normative relationships among them (or "rational

---

Y. Shoham (✉)
Computer Science Department, Stanford University, Stanford, CA, USA
e-mail: shoham@stanford.edu

Y. Shoham
Microsoft Israel R&D Center, Herzliya Pituach, Herzliya, Israel

balance", a term due to N. Nilsson). One category of metal state includes what might be called *information attitudes*, capturing agents' assessment of whether this or that fact holds true. Notions such as knowledge, certainty and belief fall into this category, and in the past quarter of a century, logics of knowledge and belief—static as well as dynamic—have become an industry, spanning disciplines as diverse as philosophy, computer science and economics (cf. [8, 23, 31]).

But information attitudes constitute just one facet of mental state. In contrast, *motivation attitudes* capture something about the agents' preference structure, and *action attitudes* capture his inclination towards taking certain actions. In a typical theory, the action attitudes mediate between the informational and motivational attitudes; the agent's choice of action is dictates by his wants and beliefs. Into these two broad camps fall notions such as desire, goal, intention and plan. The literature on such attitudes is in comparison quite thin, and the goal of this paper is to contribute to its thickening.

When one thinks about preference in isolation of other factors, things are relatively easy. There are certainly some interesting challenges—for example, capturing *ceteris paribus* conditions in preference (I prefer wealth to poverty all other things being equal, but I prefer being healthy and poor to being sick and rich) [7, 32]. But things become truly involved when one considers the interaction between the various types of attitude. For example, the dynamics of belief and preference are in general intertwined, with changes in beliefs leading to change in preference (and possibly vice versa). But more complex interactions are common. In a typical situation, one *intends* to take an action in service of a *goal* which was given rise to by certain *desires*, and conditional on some *beliefs* (for example, intending to drive your car to the city is motivated by your belief that there is a good show there that evening, and is inconsistent with believing that the car is not in working condition). Beside the interaction between the different attitudes (belief, intention, goal, desire), the discussion involves even more basic aspects of agency, such as action and ability.

This is a complicated picture, and so it's not surprising that progress on this has been relatively slow. But slow progress does not mean no progress. In the next section I discuss some of this prior work, by way of identifying different perspectives from which theories of intention are considered. The primary message of this paper that one particular perspective—the *database perspective*, which is developed further in later sections—is, from the standpoint of artificial intelligence, very useful in connection with the notion of intention.

The rest of the paper is organized as follows. In this next section I consider different perspectives from which one might study mental terms, and identify the database perspective as particularly useful in the context of intention. In Section 3 I sketch how this perspective drives the formal theory of intention and its interaction with belief. This sketch addresses the most basic form of intention, and in Section 4 I outline the various extensions of it that are required. The focus of this paper is not a specific logical system, but in Section 5 I very briefly give the highlights of a logical theory of belief and intention that follows the database perspective (this is

the theory developed in [17]). Throughout the paper I refer to prior work on intention, but in Section 6 I position the perspective offered here in the context of two pieces of prior work, both from AI, which I discuss in more detail (these are the work of Cohen and Levesque on the one hand, and Dean and McDermott's on the other). I conclude with some final remarks in Section 7.[1]

2 Criteria for a Theory of Intention, and the Database Perspective

When considering how to formalize intention—or any other complex natural notion—it is useful to consider up front the yardsticks by which one would evaluate the theory. These in turn are dictated more deeply by the sort of relevance one seeks for the theory.

One type is psychological relevance. This characterizes much of the philosophical literature on intention, a few examples of which are [2, 14, 27, 36, 38]. Thus, for example, in [3] Bratman speaks of "the psychological economy of planning agency"; his goal, and that of most other philosophers writing on the topic, is to shed light on the human experience, in this case on "practical reasoning" of the kind performed by resource-constrained human beings.

An alternative sort of relevance is social relevance, which typifies work in the social sciences. A clear case in connection with intention is [40], which studies the role of intention in the legal penal system.

A rather different type of relevance, and the one I focus on in this article, is what might be called *artifactual relevance*. This has typified work in computer science, and in particular in artificial intelligence (AI). In this case there is a particular artifact (usually defined abstractly in mathematical terms), whose behavior is completely specified and thus in principle understood, but for which one seeks an intuitive high-level language to describe its behavior. A good example is the use of the notion of "knowledge" to reason about distributed systems [8]. The protocol governing the distributed system is well specified, but intuitively one tends to speak about what one processor does or does not know about the other processor at any given state of the system (including, recursively, the knowledge of the other processor), and the role of the mathematical theory of knowledge is to formalize this reasoning. The primary message of this article is that a similar artifactual perspective can be useful in connection with intention.

These different perspectives are not mutually exclusive, and in fact there is a healthy cross-pollination among them. Thus for example the legal discussion in [40] is in direct dialog with the philosophical literature, and the Cohen and Levesque theory of intention [4]—to which I will return later—is directly

---

inspired by Bratman's theories, in particular [2]. Still there are important differences among the perspectives, with important implications to the role of logic and formal theories in general.

The philosophical discourse relies strongly on particularly instructive test cases. The "morning star—evening star" example [18] catalyzed discussion of cross-world identity in first-order modal logic (you may have different beliefs regarding the star seen in the morning from those regarding the star seen in the evening, even though, unbeknownst to you, they are in fact the same star—Venus). Similarly, the example of believing that you will win the lottery and coincidentally later actually winning it served to disqualify the definition of knowledge as true belief, and another example argued against defining knowledge as *justified* true belief [11].

Such "intuition pumps" have been used also in connection with intention. Most notably, the dentist example ("I intend to get a root canal operation, and I know that doing so necessarily entails experiencing excruciating pain, but I don't intend to experience pain") presents a requirement that one does not necessarily intend the (e.g., logical or known) consequences of one's intentions. For example, Cohen and Levesque's theory [4], discussed later, puts this forward as a major requirement. Similarly, the Little Nell example ("I believe that Little Nell is in danger, and therefore formulate a plan to save her, except that now that I intend to carry out the plan I no longer believe she's in danger") [21] calls attention to the need to account for the contexts of beliefs.

Such examples can be highly instructive, but the question is what role they play. In a certain philosophical tradition these examples are necessary but insufficient conditions for a theory. They are necessary in the sense that each of them is considered sufficient grounds for disqualifying a theory (namely, a theory which does not treat the example in an intuitively satisfactory manner). And they are insufficient since new examples can always be conjured up, subjecting the theory to ever-increasing demands.

This may be reasonable for a pre-formal theory, but not for a formal theory; a theorist would never win in this game. Consider knowledge, for example. The S5 logic of knowledge [8] captures well certain aspects of knowledge in idealized form, but the terms "certain" and "idealized" are important here. The logic has nothing to say about belief (as opposed to knowledge), nor about the dynamic aspects of knowledge (how it changes over time). Furthermore, even with regard to the static aspects of knowledge, it is not hard to come up with everyday counterexamples to each of its axioms. And yet, the logic proves useful to reason about certain aspects of distributed systems, and the mismatch between the properties of the modal operator K and the everyday word "know" does not get in the way, within these confines. All this changes as one switches the context. For example, if one wishes to consider cryptographic protocols, the K axiom ($Kp \land K(p \supset q) \supset Kq$, valid in any normal modal logic, and here representing logical omniscience) is blatantly inappropriate. Similarly, when one considers knowledge and belief together, axiom 5 of the logic ($\neg Kp \supset K\neg Kp$, representing negative introspection ability) seems

impossible to reconcile with any reasonable notion of belief, and hence one is forced to retreat back from the S5 system to something weaker (such as the S4.2 or S4.3 logics) [19, 37]. The upshot of all this is the following criterion for a formal theory of natural concepts: One should be explicit about the intended use of the theory, and within the scope of this intended use one should require that everyday intuition about the natural concepts be a useful guide in thinking about their formal counterparts.

Such a circumscribed criterion is a natural one from the artifactual point of view, and when adhered to rigorously it renders formal (e.g., logical) theory most useful. I believe that the criterion can be useful also from the philosophical standpoint, but I will delay further discussion of this to the end of the article.[2]

Since there are infinitely many sorts of artifacts, the artifactual perspective can be instantiated in many ways. In this article I want to explore a particular class of instantiations, one which seems useful in the context of intention. I will call this the *database perspective*; as the name suggests, it will again be very natural for a computer scientist, though perhaps less so for the philosopher, initially.

A database represents information in a specific format, and provides various services associated with this information, the most basic services being storage and retrieval. Logic can sometimes provide the epistemological theory of the database, capturing the semantics of the information stored in the database and of the operations on it.

This is one lens through which to view the Alchourrón-Gärdenfors-Makinson (AGM) theory [1], which has been extremely influential in the area of belief change in both computer science and philosophy in the past few decades. The AGM theory assumes that the information is any well-formed propositional sentence, and it concentrates on various operations on it, the central of which is revision. The revision operation adds a new sentence to the database, and (if needed) restores consistency while minimally perturbing the existing database. The force of the theory is in how it interprets "minimal perturbation." My goal here is not to dwell on the AGM theory *per se*, and in particular not to discuss its strengths and weaknesses,[3] but rather to suggest viewing it as specifying an "intelligent database." This database captures the current beliefs of the agent, and, in addition to the basic storage and retrieval operations, it ensures that the beliefs remain consistent at all times.

---

[2]In is interesting to note that the social-science perspective can occupy an interim position, by anchoring the theory in a social (rather than an engineering) artifact. Thus, in [40] Yaffe grounds his discussion of intention in the *Model Penal Code* [25]. While not an airtight specification, the legal language of the MPC attempts to be as unambiguous as natural language allows.

[3]As is well known, there are problematic aspects to the AGM theory. The shortcomings show up, for example, when one attempts to iterate the process of revision. Much work since that time has expanded on the original AGM formulation (cf. [23, 26]); I will return to this briefly in the next section.

Of course, AGM doesn't completely define this database. Certainly, as an epistemological theory, it has nothing to say about algorithmic issues. These were precisely the subject of *non-monotonic truth-maintenance systems* or *data-dependency systems* [22] developed roughly at the same time.[4] But in addition, the theory does not even completely specify the outcome of belief revision; it only constrains it, which admits a large class of specific revision operators that meet the constraint.

What might an analogous epistemological theory of "intention databases" look like? To answer this we should be explicit about why one might want such a database, and what uses it might have. One natural approach is to consider the intention database as being in the service of some planner, in particular of the sort encountered in so-called "classical" AI planning [39]. The planner posits a set of actions to be taken at various times in the future, and updates this set as it continues with its deliberations and as it learns new facts. In the philosophical parlance, these are "future-directed intentions." Of course, for the "intention database" to be meaningful it should provide services beyond mere storage and retrieval, just as the AGM theory specifies the service to be provided by the "belief database." What should these services be?

There is no unique right answer, as there is a range of services that could be useful. In the extreme, the entire planning process could be relegated to the database. This of course silly, but it illustrates the lack of a crisp boundary between the reasoning and storage components.[5] At the minimum, however, we should expect consistency maintenance. Analogously to the belief database of the AGM theory, intentions too must be kept consistent. Of course consistency here will mean something different; to see what it is we need to be more specific about the objects being represented.

3 The Belief-intention Database: A Sketch

We will consider information of the form "I intend to take action a at time t", where $a \in A$ belongs to a fixed set of atomic actions A, and $t \in \mathcal{N}$ is an integer. We'll call these *discrete atomic action intentions* (and usually drop the term 'discrete', leaving it implicit).[6]

Atomic action intentions admit many extensions, and I discuss them briefly in the next section. I will focus, however, on atomic action intentions, since

---

[4]Ideally the belief revision formulation would have served as a specification for the algorithmic systems, though in practice for the most part the strands of work were independent.

[5]At the risk of polluting an otherwise purely intellectual discussion, we mention that this lack of crispness is in fact familiar from the software industry, as increasing functionality is added to the database and relieved from the programmer.

[6]This is a good point at which to preempt a possible confusion. The temporal index might suggest that at the end of the day I advocate using temporal logic as a basis. As I explain in Section 5, in fact we advocate basing it on dynamic logic, taking action as the basic object rather than time. As is common in the literature, the temporal index is simply a convenient way of quantifying over action sequences of a given length.

they are the basic building block for the more complex constructs, and already involve nontrivial complications. The main complication is that planners typically associate pre- and postconditions with atomic actions. Absent the preconditions the action cannot be taken, and if it is taken the postconditions hold.[7] This means that the database must represent both beliefs and intentions, and this in turn suggests a variety of consistencies that must be preserved by the database:

1. Beliefs must be internally consistent.
2. Intentions must be internally consistent. Adopting a somewhat restrictive view of action, we might say the following:

    (a) At most one action can be intended for any given time moment.
    (b) If two intended actions immediately follow one another, the earlier cannot have postconditions that are inconsistent with the preconditions of the latter.

    Condition 2(b) can actually be deduced from the following requirements.
3. Intentions must be consistent with beliefs. This means that:

    (a) If you intend to take an action you cannot believe that its preconditions do not hold.[8]
    (b) If you intend to take an action, you believe that its postconditions hold.

A few remarks on these requirements are in order. Requirement 1 is no different from the requirement in the belief-change (e.g., AGM) theories. Requirement 2(b) is essentially Bratman's *consistency* requirement [3], instantiated to our setting. Requirement 3(a) is what is sometimes called *strong consistency*. A stronger version of this requirement would be that you believe that the preconditions of you intended action hold; this would be an instantiation of Bratman's *means-ends coherence* requirement [3]. But this does not seem useful from the database perspective, since only at the conclusion of planning— and sometimes not even then—are all these preconditions established. Making this stronger requirement will blur the distinction between the database and the planner it serves. One could also question the asymmetry between pre- and post-conditions, and specifically, in connection with requirement, why one must believe that the post-conditions of one's intentions. From the philosophical perspective this indeed might be debatable or at least very unclear. From the database perspective, however, it is a good fit with how planners operate. Adopting an optimistic stance, they feel free to add intended actions so long as

---

[7]Sometimes the postconditions are also conditional on facts that hold when the action is taken, for example "if the switch is ON then after the Toggle action it is OFF, and vice versa"—but we will ignore this complication.

[8]Both here and in 3b it is important to distinguish between the time of belief, and the time to which the belief refers. When an intention to act at time t_2 is added at time t_1 (with $t_1 < t_2$), then at time t_1 it is believed that right after the action is taken at t_2 its postconditions will hold.

those are consistent with current beliefs, but once they do they continue acting based on the assumption that these actions will be taken, with all that follows from it. Since we are not considering actions whose effects are uncertain or dependent on the conditions that obtain when the action is taken, so long as action is planned the planner firmly believes whatever follows from it. Finally, these requirements relate the conditions on belief and on intention, but do not reduce the latter to the former. Arguments for and against the alternative, reductionist view (called "cognitivist" by Bratman), which does reduce intention to belief, are discussed, among other places, in [3, 13, 14, 38]. The main lesson from all this is that whereas in the philosophical approach there is much agonizing over what the *right* definition is, in the artifactual (and in particular, database) approach the question is what a *useful* definition is. One could imagine different intelligent databases, each providing different services to the planner, and each one would be governed by a different logic.

The process of revision is made complex by these requirements. The revision of beliefs may trigger a revision of intentions and vice versa, potentially leading to a long cascade of changes. Note that facts that are believed because they are postconditions of a currently held intention must be annotated as such, so that if the intention is withdrawn then the belief in the postcondition can be eliminated as well.[9] And so we must consider the following, mutually-recursive operations on a database:[10]

–   Add a belief φ_t (optionally: annotated by action a_t):

    –   Add the belief φ_t to the belief database (optionally: add a_t to its annotation).
    –   If needed, restore consistency to the belief database using a suitable belief-change theory.[11]
    –   Repeat: So long as there is an intention $a_{t'}$ whose preconditions are violated, remove $a_{t'}$.

–   Remove a belief φ_t:

    –   Contract (in the sense of the belief-revision literature) the beliefs by φ_t. There is a question of what this means when this belief is in the postcondition of an intended action. One possibility is simply to disallow it.

---

[9]This begs the question of why not retain in general the source of different beliefs, to aid in the process of revision. We won't delve deeper into this, except to say that, indeed, why not, and to note that certain logical systems do keep track of such dependencies [5, 10].

[10]In the following, we only consider formulas φ referring to a unique time point, and indicate this by φ_t where t is the time point. This can be generalized to formulas referring to multiple time points, at considerable notational and other cost.

[11]I am deliberately non-committal here. There are known shortcomings to the AGM approach, and especially in our setting, with explicit representation of time, there may be both a need and opportunity to adopt a more nuanced approach to belief change, as advocated for example in [9]. However, belief-change *per se* is not the focus of this article, and so I prefer to sidestep this important topic.

- – Repeat: So long as there exists an intention whose postconditions entail φ_t, remove that intention.

- – Add an intention a_t:

  - – Add a_t to the intention database.
  - – If there is another intention b_t with $b \neq a$, remove b.[12]
  - – Add the postcondition of a_t annotated by a_t to the belief database.
  - – Contract the belief database by the negation of a_t's precondition. Here again there is a question of what this means when these preconditions are implied by the postcondition of an intended action. Again, one possibility is to simply disallow this case.

- – Remove an intention a_t:

  - – Remove a_t from the intention database.
  - – Delete from the database all beliefs annotated by a_t.[13]

These operations sweep a few issues under the rug. The first is obvious: With a set of mutually-recursive operations, there is the potential danger that the specification is ungrounded. However, upon inspection, this is not the case under the limitations we have imposed. This is essentially because, since we do not adopt the "means-ends coherence" requirement, intentions are never "forced in" by other changes, only forced out. And so, as long as the belief-revision part is well grounded, the system as a whole is.

A more subtle issue, however, concerns the so-called *frame problem* [20]. So far we have that the postconditions of an action hold only immediately after taking the action, but not at later times. Of course, most postconditions persist: After driving to San Francisco I remain there even after visiting the Golden Gate Park, going to dinner and then seeing a show. The fact "I am in SF" persists by default until some other action—such as "drive back to Palo Alto"—explicitly truncates this persistence. Much has been written about such default temporal reasoning and its logic (cf. [29]); it is surprisingly tricky business. Notwithstanding the logical difficulties, a version of it was incorporated in the Time Map Management System (or TMMS) [6], to which I'll return later. For now we will ignore default persistence in our belief-intention database, but ultimately this gaping hole must be plugged.

---

[12]Here we implicitly assume that the planner was aware of this conflict and decided on *a* anyway. In this respect we follow the tradition of the belief-change literature, which accords priority to the latest information received. We could of course consider other operations on the database, such as "attempt to add *a* at time *t*," which would only add a_t if that introduced no inconsistency. Similar comments apply to the other operations.

[13]This specification hides a certain complication. When the intention was added in the first place, it is possible that certain beliefs were eliminated as a result (see above). These beliefs must now be reinstated, unless there are independent reasons to exclude them. This is a complex a topic, and is related also to the complex topic of iterated belief revision. It is not possible to do this point full justice here, but I wanted to at least flag it.

A third issue has to do with whether the database is offline or online. The offline database is used by the planner in advance of actually doing anything. The online—or realtime—database is used by the planner before and during execution of the plan. So far our treatment of the database did not take realtime into account, even though modern-day planners interleave planning and execution. Fortunately, adding this component to the current framework is not hard. It requires the following:

– Equip the database with a clock, and associate the variable `now` with the value of the clock.
– Restrict the addition and removal of intentions to times t such that $t >$ `now`.

As discussed earlier, this adopts the perspective of what in the philosophical literature is sometimes called "future-directed intentions." That leaves open the question regarding the present. One can in addition require that there be no ambiguity regarding what is to be done `now`, that is, that at all times there exist an action a that the planner intends to take at time `now`. One could even require that the preconditions of a be believed (as opposed to future intentions whose preconditions merely need not be disbelieved). But this is already a matter of taste regarding the division of labor between the planner and the database, and also invites questions about how the database is to enforce this requirement. We will ignore it for now.

4 Beyond Atomic Action Intentions

There is no question that discrete atomic action intentions are very restrictive. There are various extensions that are natural to consider:

– Continuous rather than discrete time.
– Complex actions, in the spirit of dynamic logic [12]:
  – Sequences of actions or conditional actions ("I intend to take action a at time t, and at time $t + 1$ either b or c, depending on whether fact φ is true then").
  – Nondeterminism regarding the temporal dimension ("I intend to take action a sometime in the next week").
  – Nondeterminism regarding the very intentions ("I intend to take either action a_1 at time t_1 or action a_2 at time t_2").
– Achievement intentions:
  – Atomic achievement intentions ("I intend to make fact φ true at time t").
  – More complex achievement intentions, and hybrid action-achievement intentions.

- Teleology ("I intend to take action *a* since its postconditions satisfy an existing achievement intention" or "I intend to take action *a* since its postconditions satisfy a precondition of an existing action intention").
- Explicit representation of ability: So far what the agent is able to do or achieve is only inferable from the beliefs regarding the preconditions of actions. In practice, certainly there are things we know that we cannot achieve near term ("get to the moon" or "factor the product of two unknown large primes") without exhaustively reasoning about our various actions.
- Multiagent aspects: In a multiagent setting there is an interaction among the intentions of various agents. For example, I may intend to be in San Francisco only if I believe that my wife has a similar intention. Furthermore, I may adopt certain intentions as a commitment to other agents, and cannot rescind those without some appropriate protocol regarding the other agents ([30], for example, views single-agent intentions simply as commitment to oneself).

Except for the first extension, which (so long as actions remain discrete) for the most part presents no fundamental new challenges, the others pose significant new challenges to specifying semantics of the belief-intention database. For example, complex actions invoke the issue of "intention agglomeration" (if I intend A and I intend B, do I intend their conjunction?) [36]. Similarly, the multiagent setting calls for looking at the interaction between intention and game theoretic notions [28]. Various subsets of these extensions may nevertheless be required in different applications. However, as I hope is clear from the preceding discussion, discrete atomic action intentions already involve nontrivial elements, hence the focus in this article.

5 On Formal Syntax and Semantics

I will say relatively little about the sort of logic that the above analysis suggests. This is for two reasons. First, this is not the main focus of the paper. Second, it is precisely the focus of a separate paper, with Thomas Icard and Eric Pacuit [17]. Here are some high-level comments. They are quite terse and geared towards people familiar with the logical background. Others can safely ignore this section.

- Since the force of the theory is in the dynamics of intention and belief, the logic must account for the dynamics. If one wants to incorporate the realtime perspective, two natural options are temporal logic (with time as basic, whether linear or branching or otherwise) and dynamic logic (with actions as basic); both naturally incorporate the notion of `now`. Since our focus is on sentences regarding action, dynamic logic seems a natural choice.
- Our basic semantic model is thus a tree of actions separated by states. This is similar to the model studied in [24].

- Of course even within the dynamic logic setting we can employ temporal operators, which amount to quantifying over sequences of actions. This is done for example in [4].
- Belief is captured by a modal operator. Beliefs that depend on having a adopted an intention a are annotated by that intention B^a. Our beliefs play a role roughly analogous to strong beliefs in [33] and our annotated beliefs play a role roughly analogous to the weak beliefs there, but important differences hold.
- Despite its limitations, we initially adopt the AGM model; whatever concerns one has about the AGM setting, they are not exacerbated by the interaction with intention, and it is convenient to start with a well understood model. This means in particular, that the accessibility relation for belief is a total preorder, and that the static properties of belief are captured by the KD45 logic (this follows the knowledge-and-belief model of [19], among others). One needs, however, to modify the setting to capture the annotation of some beliefs by intention, as discussed earlier.[14]
- In the simple setting we defined, with only atomic action intentions, the accessibility relation for intention is the partial order defined by set inclusion; a world with the set A of intentions is preferred to a world with the set B of intention if $A \subset B$. It is not at all trivial to generalize this to the more complex settings.
- The constraints listed above suggest natural constraints between the two accessibility relations.

This suggests an extension of Dynamic Epistemic Logic (DEL) [35] to include intention (similar but also different from its extension in [27]). Since we are explicitly considering belief rather than knowledge we might replace 'epistemic' by 'doxastic,' resulting in Dynamic Doxastic-Intentional Logic (DDIL), if that's not to much of mouthful. Again, for more details see [17].

6 Past Work in AI: From Cohen and Levesque to Dean and McDermott

There is much relevant prior work, and I discussed some of it in previous sections. But the viewpoint presented in this paper can be given an additional perspective by relating it to two separate pieces of work, both in AI. The first is Cohen and Levesque's seminal [4], which spawned a string of papers on formalizing intention and related concepts in logic (for a survey of the literature see [34]; there has certainly been additional work since that survey, including [15, 27]). The second is Dean and McDermott's [6], which presented the concept of a Time Map Management System (or TMMS). Briefly, the first shares our logical approach but not our database perspective, and the second

---

[14]We emphasize though that this is a choice of convenience; as was discussed earlier, there is both a need and an opportunity to adopt a more nuanced approach to belief revision.

our database perspective but not our logical approach. In more detail the relation is as follows.

Cohen and Levesque's logic is based on a semantic model of action sequences, and uses dynamic logic as well as temporal logic operators to reason about it. To reason about mental state, it (roughly speaking) introduces three modal operators—BEL, GOAL and INTEND—for holding a belief, having a goal, and having an intention (both action and achievement). The axioms relating these concepts are involved, though perhaps inevitably so (recall Einstein's maxim that every theory should be as simple as possible but no simpler). Some features of the theory are intuitively plausible. Examples include the primary intuition guiding the paper (as captured in the title of their paper), namely that intentions are goals that tend to persist in time, as well as avoiding the dentist pitfall. Others features are less intuitive; for example, in the theory, the sentence BEL$\varphi \supset$ GOALφ is valid (though, in fairness, Cohen and Levesque are well aware of it and offer some comfort).

The perspective here shares with Cohen and Levesque many elements, both conceptual and technical, but it also diverges from them on both fronts. Conceptually, it shares the commitment to a logical theory that is informed by both philosophy and AI, but differs on how it draws on the two disciplines. They are inspired by Bratman's philosophical theory [2] and explicitly set out to capture some of the intuitions it provides. In contrast, I advocate a strict database perspective, which places crisper criteria regarding the terms that must be included in the logic and the properties that must enforced.

The discussion of technical similarities and differences is best done in the context of a specific technical proposal (e.g., [17]. However, even the elements provided here point to both similarities and differences. Certainly, we share with Cohen and Levesque the adoption of a modal operator for intention, the belief operator, and the fact that the two interact. We also find dynamic-logic operators to be convenient ways of talking about models. However, there are also substantial differences between our technical approaches. One of them concerns the models in which formulas are interpreted—they take models to be linear sequences of actions, whereas we take them to be branching models (or, action trees). The most important difference, however, is the temporal extent of a goal (and therefore also of an intention). In [4] it is left unspecified ("I intend to be in San Francisco") but is intuitively understood to be existentially quantified ("I intend to be in San Francisco sometime in the future"). I think this is an awkward choice; even on Bratman's philosophical perspective, an intention forces action. But an intention that is not anchored in time does not in general force action, as any parent of a teenager can attest. It seems to me much healthier to define time-based intentions first, and then consider various quantifications over the temporal dimension. Even then I'm not convinced that "sometime in the future" will be as useful as "by the end of next week" or even "as soon as possible," but at least we will have the basis on which to consider it.

Dean and McDermott's TMMS is a temporal database designed to aid a planner. It represents facts that are true at different points in time, as

well actions of the planner. As such it is exactly the type of database we talk about here, and in fact considers services we do not. For example, it offers a mechanism to handle the frame problem; basically, once a fact is established (for example, as a postcondition of an intention), it persists until it explicitly contradicts postconditions established by future intentions. The database also also allows certain forms of disjunctive plans. The TMMS are explicitly an algorithmic artifact, and as such can avoid thorny logical problems such as default temporal persistence or the semantics of intention. Of course, we are interested in precisely these epistemological questions. So in a sense the approach advocated here can viewed as developing the logical theory underlying (different versions of) TMMSs.

7 Final Remarks

I have argued for the value in the artifactual approach to formalizing mental state, and in particular for the database perspective in connection with intention. One could argue that this approach, while perhaps useful for some applications, does not shed light on core philosophical issues. I actually believe that the pragmatic approach forces one to confront issues that are otherwise glossed over. Obviously many of the design decisions made here make contact with notions that came up in philosophy: consistency of intentions, coherence of intentions, intention agglomeration. Of course, the very planning context is very consistent with the discussion of practical reason in philosophy. The difference is that here these notions take on a very precise meaning. To be sure this higher resolution comes at a price, since it ensures a mismatch with some elements of the human experience. But at the same time it forces one to be clear about the ontological entities participating in the discussion (events, facts, actions), and about the processes discussed (such as planning). Thus, while my perspective is firmly rooted in AI, this article aims to be relevant to the philosophical discourse as well. At least, that's my intention.

References

1. Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contractions and revision functions. *Journal of Symbolic Logic, 50*(2), 510–530.
2. Bratman, M. (1987). *Intention, plans and practical reason*. Cambridge: Harvard University Press.
3. Bratman, M. (2009). Intention, belief, practical, theoretical. In J. Timmerman, J. Skorupski, & S. Robertson (Eds.), *Spheres of reason*. Oxford: Oxford University Press.
4. Cohen, P. R., & Levesque, H. (1990). Intention is choice with committment. *Artificial Intelligence, 42*(3), 213–261.
5. de Kleer, J. (1986). An assumption-based TMS. *Artificial Intelligence, 28*(2), 127–162.
6. Dean, T., & McDermott, D. V. (1987). Temporal data base management. *Artificial Intelligence, 32*(1), 1–55.
7. Doyle, J., & Wellman, M. P. (1994). Representing preferences as ceteris paribus comparatives. In *Proceedings of the AAAI spring symposium on decision-theoretic planning* (pp. 69–75).

8. Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1994). *Reasoning about knowledge*. Cambridge: MIT.
9. Friedman, N., & Halpern, J. Y. (1999). Modelling beliefs in dynamic systems. Part II: Revision and update. *Journal of Artificial Intelligence Research, 10*, 117–167.
10. Gabbay, D. (1996). *Labelled deductive systems*. Oxford: Clarendon.
11. Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23*, 121–123.
12. Harel, D., Kozen, D., & Tiuryn, J. (2000). *Dynamic logic*. Cambridge: MIT.
13. Harman, G. (1986). *Change in view*. Cambridge: MIT.
14. Harman, G. (1999). Practical reasoning. In *Reasoning, meaning and mind* (pp. 46–74). Oxford: Oxford University Press.
15. Herzig, A., & Longin, D. (2004). C&l intention revisited. In *Proc. KR2004*.
16. Horty, J. F., & Pollack, M. E. (2001). Evaluating new options in the context of existing plans. *Artificial Intelligence, 127*(2), 199–220.
17. Icard, T., Pacuit, E., & Shoham, Y. (2009). A dynamic logic of belief and intention, (Forthcoming).
18. Kripke, S. A. (1980). *Naming and necessity* (revised and enlarged edition). Oxford: Blackwell.
19. Lamarre, P., & Shoham, Y. (1994). Knowledge, certainty, belief, and conditionalisation (abbreviated version). In *Proceedings of KR* (pp. 415–424).
20. McCarthy, J. M., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence, 4*, 463–502.
21. McDermott, D. (1982). A temporal logic for reasoning about processes and plans. *Cognitive Science, 6*(2), 101–155.
22. McDermott, D. V. (1983). Contexts and data dependencies: A synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 237–246.
23. Peppas, P. (2007). Belief revision. In F. van Harmelen, V. Lifschitz, & B. Porter (Eds.), *Handbook of knowledge representation*. Amsterdam: Elsevier.
24. Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. In *Proceedings of the third conference on knowledge representation and reasoning*. Morgan Kaufmann.
25. Robinson, P. H., & Dubber, M. D. (2007). The American model penal code: A brief overview. *New Criminal Law Review, 10*(3), 319–341.
26. Rott, H. (2001). *Change, choice and inference: A study of belief revision and nonmonotonic reasoning. Oxford logic guides*. Oxford: Oxford University Press.
27. Roy, O. (2008). *Thinking before acting: Intentions, logic, rational choice*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam.
28. Roy, O. (2009). Intentions and interactive transformations of decision problems. *Synthese, 169*(2), 335–349.
29. Sandewall, E. J., & Shoham, Y. (1994). Nonmonotonic temporal reasoning. In D. Gabbai (Ed.), *Handbook of logic in artificial intelligence and logic programming*. Amsterdam: Elsevier.
30. Shoham, Y. (1993). Agent oriented programming. *Journal of Artificial Intelligence, 60*(1), 51–92.
31. van Benthem, J. (1997). *Exploring logical dynamics*. CSLI, Stanford University.
32. van Benthem, J., Girard, P., & Roy, O. (2008). Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic, 38*(1), 83–125.
33. van der Hoek, W., Jamroga, W., & Wooldridge, M. (2007). Towards a theory of intention revision. *Synthese, 155*(2), 265–290.
34. van der Hoek, W., & Wooldridge, M. (2003). Towards a logic of rational agency. *Logic Journal of the IGPL, 11*(2), 135–160.
35. van Ditmarsch, H., van der Hoek, W., Kooi, B. (2007). *Dynamic epistemic logic*. New York: Springer.
36. Velleman, J. D. (2008). What good is a will? In A. Leist, & H. Baumann (Eds.), *Action in context*. Berlin: de Gruyter/Mouton.
37. Voorbrak, F. (1990). Generalized Kripke models for epistemic logic. In *Proceedings conference on theoretical aspects of reasoning about knowledge* (pp. 214–228). San Francisco: Morgan Kaufmann.
38. Wallace, R. J. (2006). *Normativity and the will*. Oxford: Oxford University Press.
39. Weld, D. S. (1999). Recent advances in AI planning. *AI Magazine, 20*, 93–123.
40. Yaffe, G. (2004). Trying, intending, and attempted crimes. *Philosophical Topics, 32*(1–2).