On Default Handling: Consistency Before and After

J. Drapkin M. Miller D. Perlis

Computer Science Department
University of Maryland
College Park, MD 20742
December 14, 1986

Abstract: In commonsense reasoning it is important to be able to handle conflicting data. We discuss this issue specifically in the context of default reasoning. We contrast two choices: either to constantly monitor the reasoning system in an effort to preserve consistency, or to allow inconsistencies to arise and *then* (try to) restore a semblance of order. That these are computationally virtually the same is granted; but there are other rather important distinctions between them bearing on default reasoning.

I. **Introduction**

Much of artificial intelligence research is carried on with the (often unspoken) notion that systems must be kept "consistent" at all costs. Recently, however, there has been recognition that consistency may be unrealistic for many purposes. Levesque [84], Konolige [85], and Fagin and Halpern [85], are among those who have explored formal approaches to inconsistent reasoning systems. When dealing with commonsense reasoning, it is especially important to be able to handle conflicting data. Here we discuss this issue specifically in the context of default reasoning.

Our discussion centers on two contrasting choices: either to constantly monitor the reasoning system in an effort to preserve consistency, in the sense of (trying to) *guarantee beforehand* that a proposed new conclusion will not upset consistency; or to allow inconsistencies to arise and *then* (try to) restore a semblance of order. That these are computationally virtually the same is granted; but there are other rather important distinctions between them bearing on default reasoning.

II. **The Example**

A standard example of default reasoning, and one that we will use to illustrate our arguments, is as follows:

> Bird(x) $\tilde{\rightarrow}$ Flies(x) [axiom]
>
> Bird(Tweety) [observation or axiom]
>
> Flies(Tweety) [conclusion]

where the $\tilde{\rightarrow}$ indicates a qualified (default) rule; that is, Flies(x) is to be concluded if Bird(x), *unless* it is inappropriate to do so, that is, unless something is known that blocks the conclusion Flies(x). We call this condition an unless-condition. Typically, the unless-condition involves a consistency test, to wit, that the proposed conclusion be consistent with what is already known. We will take this to be understood except when specified otherwise.

Now suppose the following are learned:

> Ostrich(x) $\rightarrow$ ¬Flies(x)
>
> Ostrich(Tweety)

No longer is the wff Flies(Tweety) derivable (in the augmented system of axioms) since now the $\tilde{\rightarrow}$ will, on the basis of the new information, block its normal conclusion of Flies(x).

We will analyze this example in detail in what follows. Our hope is to show that consistency is preserved here in only a very limited sense, and that insight into the process of default reasoning is gained by letting potential inconsistencies be explicit.

Throughout the paper, our focus will be on reasoning systems viewed as having reasoning-lifetimes, i.e., instead of simply providing one answer to one query, they continue to reason and revise their beliefs over time.

III. **Background on the Treatment of Consistency**

Perhaps the principal reason that consistency is generally regarded as a *sine qua non* in artificial intelligence, is that any logic-based system that admits inconsistency in its axioms will also admit all formulas as theorems. However, it must be pointed out that this does *not* mean all formulas will surface as the system "runs". For instance, the algorithm used to find theorems may be far from complete. Moreover, lack of completeness need not be a disadvantage, especially in a commonsense system, as opposed to a theorem-prover that may put a premium on depth rather than breadth.

Nonetheless, if an inconsistency does surface, even an incomplete reasoner is likely to suffer difficulties, unless special care is taken, so the ideal of consistency has a certain appeal. On the other hand, consistency checking typically will lead to exponential explosion, or even undecidability. Thus when it is desired to know that the addition of some new fact C will not introduce inconsistencies within an existing framework, it must be shown that C is consistent with all logical consequences of all that is currently known.

With defaults, there is an even greater level of concern. Here, if it is to be truly automatic, the system itself must check consistency, and then later resolve any inconsistencies that might arise. Thus the best-known approaches to default reasoning employ some kind of formal consistency checks; these are precisely the unless-conditions mentioned above.

A second reason to shun inconsistent systems is that they have no models, hence (so one may argue), there is no true interpretation for them. If so, then it is hard to assess what they are about. However, if we view such a system as an object of study, then we can form our own consistent meta-theory about it. This in fact seems to be what is typically done, for instance in the formal study of belief logics.

The theorem-proving-oriented approaches (see McCarthy [80], McDermott and Doyle [80], and Reiter [80]) are especially prey to these objections vis-a-vis inconsistency. In the notation of McDermott and Doyle, the default rule for our example would be a single axiom, namely,

$$\{ \text{Bird}(x) \text{ and Consis}[A \ \& \ \text{Flies}(x)] \} \rightarrow \text{Flies}(x)$$

where A consists of what is already known (and "trusted"). Reiter's approach employs instead a rule of inference of much the same form. McCarthy has a significantly different approach (circumscription) involving an axiom schema. The real-time approach of Drapkin, Miller, and Perlis [86a] instead models the step-by-step reasoning of the system, where the consistency proofs (which at the very least involve many steps) are replaced with simple checks for the presence of direct contradiction. That is, an unless-condition in that approach takes the form

$$\{ \text{Bird}(x) \text{ and } \neg\text{Proven}(\neg\text{Flies}(x) \} \rightarrow \text{Flies}(x)$$

where Proven refers to wffs already established in the ongoing process of reasoning. We will discuss these differences later.

IV. **The Idea: Consistency-Before Equals Contradiction-After (sometimes)**

Our starting point is the commonplace observation that checking for consistency is the same thing as checking for the absence of contradiction. While this is rather trivial, it has interesting consequences. Firstly, it shows us that the effort spent in invoking the unless-condition in a default rule is the same as that in simply ignoring the unless-condition, drawing the "bald" conclusion (Flies(Tweety)) and *then* looking for a possible contradiction. This may sound like heresy, in that it openly invites inconsistency, not to mention outright falsehood: it is not true that all birds fly.

However, we contend that in a lifetime-reasoner that revises its beliefs rather than simply producing the "right" beliefs when queried, the appearance of consistency via the unless-condition is misleading. Specifically, in order to recognize that an old belief (Flies(Tweety)) is now to be replaced by a new one (¬Flies(Tweety)), a momentary inconsistency is noticed. This phenomenon is not dealt with in much of the literature on defaults, simply because the context has usually been that of single-answer systems as opposed to ones that can revise old beliefs. Indeed, we will argue that there is an advantage in waiting for inconsistencies rather than keeping them out beforehand.

Let us continue our discussion in terms of our example. Suppose as before that the conclusion Flies(Tweety) is drawn on the basis of an unless-condition (nothing already known contradicts Flies(Tweety)). When it is later learned that Tweety does not fly, the default rule *no longer* can draw this conclusion. However, this should not obscure the fact that it *has already* drawn that conclusion. We must face an issue not usually addressed in formal studies of default reasoning: when beliefs are revised, old beliefs do not magically vanish in a puff of smoke. Rather, some mechanism must decide that a revision is called for. We cannot simply disregard *all* previous beliefs, for that would amount to losing the many facts that *are* still trustworthy. Distinctions, then, must be drawn between those old beliefs that are still workable and those that are not. This of course is the frame problem (Doyle's TMS [79] solves part of this). Thus belief revision is not a simple matter, even when unless-conditions have seemingly kept things consistent.

In fact, the consistency provided by unless-conditions (that attempt to ensure that a new conclusion will not conflict with other facts) is only local. That is, consistency is maintained only among the "current" conclusions, the ones deducible under the current state of the system -- but these conclusions should be seen as residing within a larger context of the system's earlier reasoning. Here consistency is not achieved without major additional work. Even then, it is consistency-restoration rather than consistency-preservation. There is therefore an unavoidable and major component to consistency that necessarily occurs *after* conclusions are drawn. We contend that there is little advantage (and real disadvantage) to performing *beforehand* consistency tests at all. This is our main point, which we will amplify in the following section.

V. **The Example: Beforehand vs. Afterwords**

Suppose as in our original example, it has been concluded that Flies(Tweety), and later it is learned that Ostrich(Tweety). Now, if we are dealing with a lifetime-reasoner, the fact that this new axiom has been "learned" does not in itself say that all old conclusions are to be discarded. Rather, it must be inferred that now, because ¬Flies(Tweety) is deducible, and because this conflicts with the old conclusion (Flies(Tweety)), then one of the two must be surrendered. It is true that in some sense Flies(Tweety) is no longer among the "current" beliefs, in that taking only the original axioms (not their consequences) plus the new axiom (Ostrich(Tweety)), leaves us with a system in which Flies(Tweety) indeed is not deducible. However, as we pointed out above, it is folly for a system to disregard all its old conclusions, even all its old default conclusions. Hence it must pick and choose among them, on the basis of conflicts with new conclusions. That is, it once again must perform consistency checks *after* those conclusions are drawn, not to sanction the new conclusions this time, but rather to "clean-up" its world-view regarding these old and new conclusions. The difficulty we are addressing can be seen as that of deciding what is to be taken as A (recall A is what is known and trusted), when the system evolves into a new world-view. Below we illustrate in more depth why this is a problem.

The "consistency beforehand" approach, then, is that of verifying that A+C is consistent (where C is the conclusion that one wishes to draw) *before* C is concluded. There are several difficulties with this approach. Perhaps the most salient one is simply this: Consistency checking, even at its best, is a slow process. Since the whole point of such tests in default reasoning is to sanction default conclusions, i.e., statements that typically are true and should be sanctioned, then why spend time during which the eventual conclusion is held up? Why not instead let it go through, and in those rare cases where it was a mistake, *then* correct it? We have already addressed the argument that at times this temporarily will allow an inconsistency into the system, by noting that inconsistencies will unavoidably arise anyway in a system with a history (i.e., a lifetime-reasoner). Our suggestion then is to "push ahead" with the typical conclusion, since we expect it to turn out that way; that is, usually the beforehand consistency test will turn out to have been unnecessary, a no-op. We note that the time spent checking for contradiction *after* the rule is applied can be done *in parallel* while the system proceeds to use its default conclusion to continue its reasoning.

Consider the case of an ostrich, Tweety, known not to fly. In the afterwords approach, on the basis of being a bird, it would be concluded that Tweety can fly. While this conclusion is being checked subsequently for consistency, it may also be used to conclude that Tweety nests in trees. When eventually the contradiction (fly vs. not fly) is found, the damage has been done in the form of erroneous side-effects, and cleaning-up is called for. On the other hand, the beforehand approach is no better off, for it may have been concluded that Tweety, an ostrich, is a fast runner and consequently is safe from wolves. When it is discovered later that Tweety has a broken leg, and hence *cannot* run (*further* consistency checking is required to notice this clash), the same kind of undesirable side-effects have occurred as in the afterwards approach. Thus consistency checking afterwards and cleaning-up are necessary parts of reasoning in the commonsense world; no amount of consistency checking beforehand will obviate that.

It may be of interest to mention the amount of time that must be spent to determine if the addition of an assertion is consistent. With the approaches of McDermott and Doyle and of Reiter, the time factor is lethal, for *all logical consequences* of the current set of beliefs must be checked, an undecidable matter in general; in effect, beforehand checks result in infinite slowdown. With McCarthy's approach of circumscription the situation is better, but potentially unbounded time is needed to find and draw consequences of substitution formulas for the predicates to be circumscribed. Even with the Drapkin, Miller, and Perlis method, the beforehand checks result in a slowdown of about a factor of three. Now, these rough measures will still apply to a consistency-afterwords approach. But then the default conclusion has already been installed and can be used while the afterwords test is still going on.

Thus the use of typicality in the afterwords approach is not kept waiting by a potentially lengthy process (theorem-proving). Indeed, this seems to be in the spirit of default reasoning, in which a default rule allows a short-cut to a likely conclusion, without full verification. After all, one *might* be able to prove rigorously that Tweety flies, either by calculations (based on extensive information about Tweety), or more likely, by seeking new data. Either of these may take less time in general than carrying out an actual consistency proof.

The second difficulty we wish to point out with the beforehand approach is that it amounts to unnecessarily dividing forces, in effect multiplying the work. For instead of a single (though difficult) cleaning-up action *afterwards*, now there will be two: one before and one after. Since these to a large extent repeat one another, there appears to be some wasted effort.

The third disadvantage we can argue only tenuously, although we think it is a significant point. It is that often it is very important to notice conflicts with already accepted beliefs, so that one may re-evaluate them. But to do this, a system must *not* block any potentially new belief that may conflict with old beliefs. In a sense this is the flip-side of an earlier point about prior beliefs, but now concerning prior non-default beliefs. For instance, suppose a system has the default rule that for any day d, the sun will rise on next(d) (i.e., tomorrow), other things being equal. If it also has been told that today is the last day of the world, then it will not, in the normal "beforehand" approach to defaults, conclude that the sun will rise the next day. On the other hand, it seems highly appropriate that it should notice a conflict and decide that something is wrong (perhaps its belief that the world is about to end is in error), rather than blindly assume that everything it happens to think is necessarily the case. That is, it can be very healthy to take the attitude of re-weighing beliefs in the light of new evidence, even if this new evidence itself is controversial and default-based.

This example can be appreciated more when interpreted as follows: Suppose the belief that the world is about to end results from an assertion made by another agent, taken to be true on the basis of a default rule that others should be believed unless there is counter-evidence. Then how are we to resolve the conflict between the two default rules? Surely we must allow the two conclusions to fight it out between themselves, on the basis of commonsense, i.e., world knowledge. We could (as suggested by Nutter [83]) mark default conclusions as tentative ("presumably true"), thereby avoiding an outright inconsistency; but then we may well find that over time nearly all the system's beliefs become tentative, in which case nothing has been gained. Elsewhere (Drapkin, Miller, and Perlis [86b]) we argue this in greater detail, as an outcome of the frame problem.

We now present a sketch of our view of how things "should" be done, i.e., the afterwords approach. Suppose, instead of checking for the consistency of A + C before C is concluded, the system simply concludes C outright. That is, instead of using the *default* rule,

$$\text{Bird(x)} \ \tilde{\rightarrow} \ \text{Flies(x)}$$

we envision a "brute force" rule such as

$$\text{Bird(x)} \rightarrow \text{Flies(x)} \ .$$

That is, given Bird(x), we would *immediately* conclude Flies(x). We would *then* have to check if any contradictions arose. If so, an appropriate algorithm would be employed to resolve the inconsistency.

To forestall an objection: A "squelch" device will be needed to keep the system from constantly re-deducing and resolving the same inconsistencies. But this is required also in the beforehand approach, where the unless-condition itself is a constant squelch used to block the given rule from firing. An added advantage of a squelch mechanism afterwords is that it can offset the unpleasant taste some may find in the "false" axiom Bird(x) → Flies(x), for we can write instead a rule of inference: from Bird(x) infer Flies(x), employ a squelch mechanism to keep this from firing when evidence warrants, and have another axiom Typically(Bird(x) → Flies(x)) to allow the system to know the "real" situation. Clearly, the successful use of such a device requires making design decisions. In the beforehand approach, this would amount to the usual consistency kinds of checks; in the afterwords case this could be used as well, although we are experimenting with real-time approaches to this as well. See (Drapkin, Miller, and Perlis [86a]).

VI. **Conclusions**

In summary, dealing with conflicting information may be the real heart of default reasoning. If so, then the consistency tests made beforehand defeat the purpose. We have indicated various difficulties with that approach. Tests made afterwards do not solve this, but they do allow it to be addressed. Some of the ramifications and advantages of this have been explored. While our suggestions may appear to clash with general preconceptions about the nature of reasoning, we think that there is much to be gained by experimentation with these ideas.

Bibliography

(1) Doyle, J. [1979] A truth maintenance system, AIJ 12.

(2) Drapkin, J., Miller, M., and Perlis, D. [1986a] A memory model for real-time commonsense reasoning. Submitted for publication.

(3) Drapkin, J., Miller, M., and Perlis, D. [1986b] The frame problem and natural kinds. Working paper.

(4) Fagin, R. and Halpern, J. [1985] Belief, awareness, and limited reasoning: preliminary report, IJCAI-85.

(5) Konolige, K. [1984] Belief and incompleteness. SRI Tech Note 319.

(6) Levesque, H. [1984] A logic of implicit and explicit belief, AAAI-84.

(7) McCarthy, J. [1980] Circumscription--a form of non-monotonic reasoning, AIJ 13.

(8) McDermott, D. and Doyle, J. Non-monotonic logic I, AIJ 13.

(9) Nutter, J. T. [1983] Default reasoning using monotonic logic: a modest proposal, AAAI-83.

(10) Reiter, R. [1980] A logic for default reasoning, AIJ 13.