Defaults Denied

Michael Miller*, Donald Perlis† and Khemdut Purang‡

mm@i-a-i.com, perlis@cs.umd.edu, kpurang@cs.umd.edu

June 10, 1996

Abstract

We take a tour of various themes in default reasoning, examining new ideas as well as those of Brachman, Delgrande, Poole, and Schlechta. An underlying issue is that of stating that a potential default principle is not appropriate. We see this arise most dramatically as a problem in an attempt to formalize what are often loosely called "prototypes", although it also arises in other formal approaches to default reasoning. Some formalisms in the literature provide solutions but not without costs. We propose a formalism that appears to avoid these costs; it can be seen as a step toward a population-based set-theoretic modification of these approaches, that may ultimately provide a closer tie to recent work on statistical (quantitative) foundations of (qualitative) defaults([1]). Our analysis in particular indicates the need to resolve a conflation between use and mention in many default formalisms. Our treatment proposes such a resolution, and also explores the use of sets toward a more population-based notion of default.

*Intelligent Automation, Inc., Rockville MD 20850

†Institute for Advanced Computer Studies and Department of Computer Science, University of Maryland, College Park, MD 20742

‡Department of Computer Science, University of Maryland, College Park, MD 20742

Contents

1 Introduction

Default principles (which we will usually refer to simply as *defaults*) are generally expressed as typicality statements of the form "typically X's are Y's", where X and Y are general classes of objects. Much effort in artificial intelligence has been directed toward the development of logic-based approaches for interpreting, representing, and reasoning with and about typicality statements. A variety of formalisms have been proposed including, but not limited to, Delgrande's conditional logic[4], [5] McDermott and Doyle's NML[21], Reiter's Default Logic (DL)[36], McCarthy's circumscription[17], Poole's hypothetical reasoning framework[32], and Schlechta's treatment of defaults as generalized quantifiers[37].[1] Each of these formalisms treats typicality statements—which henceforth we denote $X \overset{typ}{\rightarrow} Y$ without intending any bias toward a particular formalism—somewhat differently in its interpretation of typicality, in technique, and in its success in modeling the whole of default reasoning.

In this paper we take a tour of various themes in default reasoning. The tour traces an odyssey that we have taken, beginning with a representational difficulty that arose in our attempt to formalize what are loosely called "prototypes," and which to avoid terminological confusion we will define below as *individual-inference-stand-ins*. This leads to related difficulties in more traditional default formalisms. An underlying issue is that of stating that a potential default principle is *not* appropriate. While we see this arise most dramatically in our attempt to formalize individual-inference-stand-ins—which may differ conceptually from the more standard "typicality-based" default treatments—the problem does surface in these other approaches as well.

We first noticed the need for denials of defaults in exploring the issue of "range defaults" [23]. In that and other papers [24, 25] we have continued to investigate denials. Several other researchers (Delgrande [4, 5], Poole [32], Schlechta [37]) have independently studied default denials.

In section 3 we present our treatment of reasoning based on individual-inference-stand-ins, and then show in section 4 that this treatment is subject to substantial difficulties, specifically regarding the representation of the denial of defaults (i.e., asserting that a given default principle does *not* hold). We then point out similar difficulties in the more traditional default formalisms, which Delgrande, Poole, and Schlechta have also noted and addressed. In section 5 we discuss and evaluate the Delgrande, Poole, and Schlechta proposals. In section 6 we analyze the problem in more detail and suggest that it arises when a formalism promotes the conflation of two distinct components of default reasoning: the inferential *use* of a default principle and its descriptive *mention*. We then present our own solution, which is to separate use from mention. We show that the application of our solution to traditional formalisms requires only the addition of a single first-order predicate (in a quotational language), thereby achieving both use and

[1] See [9] for a good review of several of the logic-based formalisms mentioned as well as other non-logic-based formalisms.

mention capacities, in contradistinction to other approaches that undertake massive revision of traditional NMR in order to formulate denials.

2 Terminology

Despite their differences most, if not all, of the standard default formalisms share the common view that a typicality statement is a *rule of thumb* that can be applied (in a way highly dependent on the formalism) to objects represented in the reasoner's ontology. Thus we use the $X \stackrel{typ}{\rightarrow} Y$ notation in a (hopefully) neutral manner, which can be regarded as applying equally to default logic, circumscription, and so on.

In order to achieve clarity in our discussion, we offer some additional terminological guidelines related to what appears to be in common informal usage, as well perhaps in Cognitive Psychology. There are many terms that are used for similar and not always clearly-delimited notions: archetype, prototype, generic element, default concept, exemplar, typical element, etc. At the risk of adding to the confusion, we will provide yet additional terms, but with (we hope) clear definitions. Some of these terms will be used in a later section in an (ultimately abortive but revealing) attempt toward a new default formalization based on these terms, that will in turn lead to a much improved formal variant on traditional default approaches.

A *stand-in* SI for a class C is a conceptual entity (in a reasoning agent G) that "comes-to-mind" in G's thinking when class C is mentioned. This does *not* require G to believe that most members (or indeed any members) of C are much like SI at all. As an example, suppose G invariably thinks of a vivid blue bird whenever birds are mentioned, even though there is not a preponderance of vivid blue birds and even if G knows this. G may wisely refrain from drawing conclusions about birds based on characteristics of his stand-in SI, even while continuing to have thoughts of SI when birds are mentioned.

Even when a stand-in is regarded as very like most members of C, it may fail to be used to draw conclusions about individuals, and for good reason. For instance, in deciding grades in a certain course, a teacher may from past experience have excellent reason to believe that at least 85% of the students will receive the grade of B. Yet he is not likely to use this belief to form and employ a default principle to the effect that any particular student will receive a B; rather, he examines each student's record to see what grade has been earned. In fact, if the teacher were to use such a default, it may well lead to a version of the Lottery Paradox [15, 20, 10, 35], since he may also have a strong belief that at least, say, 5% of the students will receive an A: then the default could lead him to guess that each individual student gets a B, which conflicts with the belief about some students receiving A's.

Thus the decision to *make inferences about individuals* from information about what most members of a class are like, is not the same thing as having strong beliefs about what most members are like.

In some situations, at least, outcomes (e.g., grades) are of a kind that calls for finding out more about individuals rather than basing decisions on initial, very limited information. Thus reasoning *about* possible defaults and their appropriateness or inappropriateness becomes very important. If forced to guess what a student's grade will be, the teacher might well guess B, but yet will refrain from treating this as the actual grade.[2]

Such stand-ins would appear to figure quite a bit in human thought. Further examples are "the girl next door" (as a stand-in for the class of girls) and "the friendly neighborhood cop" (as a stand-in for the class of police officers). These may be fictitious entities altogether, or largely or fully based on real persons. And G need make no assumptions about girls and police based on such stand-ins.

We thus distinguish a special kind of stand-in SI that G takes to be very like a preponderance of the members of C; and from which G is disposed to draw conclusions about a given member M of C, namely that M shares with SI any properties of interest, unless there is reason to refrain from this. We will call such an SI an *individual-inference-stand-in*.[3] As an example, G may in fact believe a preponderance of birds are blue, and may also comfortably draw the (defeasible) conclusion that a particular bird of interest is blue from this.

Even if G employs an individual-inference-stand-in for C, there may or may not exist an actual member of C that is similar to a preponderance of the members of C in all respects of interest. An individual-inference-stand-in is a conceptual entity in G's thinking.

To gather the above into a definition: An *individual-inference-stand-in* of a class C is a conceptual entity IISI that is associated by a reasoning agent G with members of the class, reflecting a tendency on G's part to recall and reason with IISI when class C *or a member thereof* is under consideration. In particular, G may treat such a member as similar to IISI when little detail about that member is available.

The above terminological distinction between stand-ins in general and individual-inference-stand-ins allows us to make a further distinction that will inform the rest of our paper: G may observe that his stand-in SI is in fact *not* typical of class C; in so observing, G is not using SI to draw conclusions about any particular individual member of C, but rather is *mentioning* a fact *about* how C and SI are related. Similarly, G may observe (mention to himself) that his stand-in SI *is* typical of C and furthermore appropriate for use as an individual-inference-stand-in and may then proceed to actually use this to conclude of member M of C that it has property P in common with SI.

[2] This may well be thought of as having one's default and eating it too, a delicate default meta-reasoning that surely is part of commonsense. One aspect of this is a principal focus of this paper: how to represent the belief that a given default is inappropriate for use despite perhaps correctly reflecting what most of a class is like.

[3] This may bear an interesting relationship to the notion(s) that exemplars and prototypes correspond to in Cognitive Psychology [8, 38, 22]; but we do not wish to take sides among the various competing theories there, hence we have adopted our own neutral term.

The distinction we wish to emphasize is that G can both *use* and *mention* individual-inference-stand-ins, and that these are not the same activities. This is most evident when G mentions that a stand-in is *not* typical, and this is where default denials will enter. However, even mentioning that a stand-in *is* typical is not the same thing as using it to draw a default conclusion about a particular member of C, nor even the same thing as believing one should so use it. This rather obvious distinction will play a central role in what follows. The one (mention) is a statistical assertion relating the entire class C to a stand-in, and the other is a disposition to reason to a conclusion about an individual member of C. This will be treated in detail in section 6.

3 Formalizing Individual-Inference-Stand-Ins

We now try to formalize individual-inference-stand-ins, which will turn out to be a very different way to view typicality (and hence defaults) from the usual formal treatments. In fact, it will ultimately be seen not to work, but for reasons that will then bear significantly on all default formalisms.

For each general class of interest, C, we designate a special fictitious[4] element of C as its individual-inference-stand-in: $iisi(C)$. It seems natural to consider that this fictitious object's having certain properties is what "encodes" the defaults about the concept that it represents. So for example for a given agent G, $iisi(Tree)$ may have the property "has leaves" (which might be thought of as amounting to G's default that "typically trees have leaves"): $Tree \stackrel{typ}{\rightarrow} Has\_Leaves$. G's reliance on the properties of $iisi(Tree)$ to (defeasibly) decide about the status of a particular real tree (e.g., that it too has leaves) amounts to G's predisposition to use that default.

Here we extend a first-order language to include representations of these mental notions in the form of constant symbols called *iisi*-constants. For each formula Φ with one free variable we write $iisi_\Phi$ to denote the individual-inference-stand-in for the class of Φ-objects. As reified "objects of thought" iisi-constants have properties and are subject to manipulation in the reasoning process. That they are not "real" objects forces us to treat them somewhat judiciously and not altogether like actual objects, but they are objects to reason about nonetheless.

Enriching a formal reasoner's ontology to accommodate mental notions or concepts in general is not a new idea, nor is it new to AI ([33],[34] and [16] are examples). Indeed, Brachman[3] discusses the very idea of reifying prototypical mental notions to represent defaults in frame-like inheritance networks, yet no one seems to have carried this out in a logic-based approach to see what, if any, advantages and/or shortcomings might arise. One of Brachman's objections to the treatment of typicality in frame-like

[4] There is an extensive and controversial literature on non-existent objects. See [14] for a good review. We will not be concerned with these issues here.

approaches is that the interpretation of the typicality concepts themselves is open to confusion and debate. There is no such confusion in the formalism we present here; a *iisi*-constant is, indeed, a "prototypical *individual* that somehow typifies the kind" ([3] p. 89). We will find that fictitious individuals encounter a severe problem, even inconsistencies, when exposed to the rigors of formalism (Section 4). In the process, however, we will learn a valuable lesson that will be applied to standard default formalisms in later sections.

3.1 A formal approach

3.1.1 *iisi*-constants

The *iisi*-constant formalism is based on a standard first-order theory $\mathbf{T}$ with language $\mathcal{L}$. We will describe an extension $\mathbf{T}^{iisi}$ of $\mathbf{T}$ for reasoning with *iisi*-constants. By way of notational convention, let Ψ, Φ, and various subscripted Φ_i's stand for formulae of $\mathcal{L}$ containing one free variable. We extend $\mathcal{L}$ to include one new constant symbol $iisi_\Phi$ for each such Φ in $\mathcal{L}$. We will call this extended language $\mathcal{L}^{iisi}$. The symbol a will be used to stand for a closed term in $\mathcal{L}^{iisi}$.

As alluded to above, the intended interpretation of the constant $iisi_\Phi$ is the reasoner's notion of the typical-Φ-object; for example, $iisi_{\mathrm{Bird(x)}}$ denotes the typical-bird-object. (Note: From here on, for convenience, we drop the '(x)' from the *iisi*-constant notation, and will write, for instance, $iisi_{\mathrm{Bird}}$ instead of $iisi_{\mathrm{Bird(x)}}$.) Defaults are encoded as expressions of the form $\Phi(iisi_\Psi)$, which can be read "the typical-Ψ-object has property Φ" or "the typical-Ψ-object is a Φ." For instance, the default that "the typical bird flies" (or "typically birds fly") is written $Flies(iisi_{\mathrm{Bird}})$.

Sometimes, when discussing the *iisi*-constant approach to defaults, we will revert to the neutral '$\Psi \xrightarrow{typ} \Phi$' notation instead of '$\Phi(iisi_\Psi)$'—notice the transpositioning of 'Φ' and 'Ψ'—because it is closer to the traditional way of depicting and reading defaults. Thus in the *iisi*-constant approach $\xrightarrow{typ}$ is *not* a new logical connective. Note that although in $\Psi \xrightarrow{typ} \Phi$ there is an *apparent* free variable,[5] the intended reading is "typically Ψ's are Φ's" with no free variable. Thus $Bird(x) \xrightarrow{typ} Flies(x)$ should be regarded as a closed sentence. Moreover $Bird(tweety) \xrightarrow{typ} Flies(tweety)$ is ill-formed; there is no substitution of terms permitted for the apparent free variable in $\xrightarrow{typ}$ expressions. It may be less confusing to write $Bird \xrightarrow{typ} Flies$ for the default principle.

In addition to the axioms of the original theory $\mathbf{T}$, the extended theory $\mathbf{T}^{iisi}$ contains one proper axiom schema:

$$\Phi(iisi_\Phi) \qquad \textbf{Schema A}$$

[5] Because Ψ and Φ have one free variable each.

which just assures that Φ applies to $(iisi_\Phi)$; e.g., that the typical bird is a bird (i.e., $Bird(iisi_{\mathrm{Bird}})$), the typical singing bird sings and is a bird (i.e., $Sings(iisi_{\mathrm{Sings}\wedge\mathrm{Bird}}) \wedge Bird(iisi_{\mathrm{Sings}\wedge\mathrm{Bird}})$), and so on.

In addition to **T**'s inference rules, $\mathbf{T}^{iisi}$ contains one default inference rule:

$$\frac{\Phi(iisi_\Psi),\ \Psi(a),\ Unknown(\neg\Phi(a))}{\Phi(a)} \qquad \textbf{Rule D}$$

which sanctions the judicious use of the encoded default appearing as the leftmost component to the antecedent of the rule. The intuition behind the rule is this: if a is a Ψ-object then assume it to be as much like the typical-Ψ-object as possible. Thus if the typical-Ψ-object has property Φ then so too should a, unless known to the contrary.

The condition $Unknown(\neg\Phi(a))$ attached to rule **D** represents a criterion that tests for the appropriate application of the default $\Phi(iisi_\Psi)$, thereby giving the formalism its nonmonotonic flavor. For our purposes we need not choose a particular implementation of $Unknown$, though several possibilities come to mind, including fix-point consistency checking in the style of Reiter's Default Logic (DL) and McDermott and Doyle's NML (both undecidable), circumscription (semi-decidable), and the negative introspection facility of step-logics (decidable)[6], [7].[6]

3.1.2 Some pleasant features

One nice feature of the *iisi*-constant approach to default reasoning is that, like circumscription, it requires no special logical connective (e.g., Delgrande's '$\Rightarrow$' connective [4], [5]) which is semantically distinguished from first-order material implication in order to write defaults. Nor are we committed to a modal operator that loosely corresponds to our $Unknown$ (e.g., the 'M' operator of McDermott and Doyle[21]).

Delgrande helps point out another nice feature of the *iisi* formalism by distinguishing the ability to reason *with* defaults from the ability to reason *about* defaults. Reasoning *with* a default is simply using it to come to a conclusion about an individual instance of the default rule. Thus we infer that Tweety flies because she is a bird, birds typically fly, and we don't know otherwise. Default formalisms are geared toward this kind of reasoning and hence have a fair amount of success with it.

Reasoning *about* defaults, by contrast, allows a reasoner to infer a new default, perhaps from some starting set of defaults.[7] This, as Delgrande observes, is in general beyond the reach of DL and NML. For example to infer that "typically birds have wings" from "typically birds fly" and "typically flying things have wings"[8] is impossible in DL. DL's defaults are themselves inference rules, not part of the logical language, so to infer a new default would require a mechanism within the logic that adds new inference

[6] In NML and circumscription, rule **D** would be written as an axiom schema rather than an inference rule.
[7] We will see even other abilities to be subsumed under *reasoning about defaults* later.
[8] Or even to infer that "typically birds have wings" from "typically birds fly" and "*all* flying things have wings".

rules to the theory. DL has no such mechanism; nor does any other well-known logic.

The *iisi*-constant theory presented here is able to model both reasoning with *and* about defaults (as can one of Delgrande's formalisms [5] and circumscription). To see this notice that the a in rule **D** is not confined to the original language $\mathcal{L}$, but rather can be any closed term in $\mathcal{L}^{iisi}$, and in particular may itself be an *iisi* constant (i.e., a constant symbol of the form $iisi_{\Phi_1}$). In the case where a is in $\mathcal{L}$, rule **D** operates, in spirit, much like a default rule in DL whereby a "real" domain object can be attributed a property by default. So, for example, $Flies(tweety)$ follows from $Flies(iisi_{\mathrm{Bird}})$, $Bird(tweety)$, and $Unknown(\neg Flies(tweety))$. In the case where a is of the form $iisi_{\Phi_1}$, defaults can be combined to create new defaults. For example $Winged(iisi_{\mathrm{Bird}})$ follows from $Flies(iisi_{\mathrm{Bird}})$, $Winged(iisi_{\mathrm{Flies}})$, and $Unknown(\neg Winged(iisi_{\mathrm{Bird}}))$.[9] More generally, the following results characterize, in part, the way defaults are inferred in $\mathbf{T}^{iisi}$.

Theorem 3.1 (Transitivity) For all monadic predicate expressions Φ_1, Φ_2, and Φ_3

$$\Phi_2(iisi_{\Phi_1}), \Phi_3(iisi_{\Phi_2}), Unknown(\neg\Phi_3(iisi_{\Phi_1})) \vdash_{T^{iisi}} \Phi_3(iisi_{\Phi_1})$$

(Note: The justification for the term "transitivity" becomes evident when the theorem is written using the $\overset{typ}{\rightarrow}$ notation as in: $\Phi_1 \overset{typ}{\rightarrow} \Phi_2,\ \Phi_2 \overset{typ}{\rightarrow} \Phi_3,\ Unknown(\neg(\Phi_1 \overset{typ}{\rightarrow} \Phi_3)) \vdash_{T^{iisi}} \Phi_1 \overset{typ}{\rightarrow} \Phi_3$.)

Proof: Using the default rule **D** where Φ is Φ_3, Ψ is Φ_2, and a is $iisi_{\Phi_1}$ the result follows immediately. $\square$

Theorem 3.2 (Compositionality) For all monadic predicate expressions Φ_1, Φ_2, and Φ_3:

(a) $\Phi_2(iisi_{\Phi_1}),\ \forall x\{\Phi_2(x) \rightarrow \Phi_3(x)\} \vdash_{T^{iisi}} \Phi_3(iisi_{\Phi_1})$

Or, in $\overset{typ}{\rightarrow}$ notation, $\Phi_1 \overset{typ}{\rightarrow} \Phi_2,\ \forall x\{\Phi_2(x) \rightarrow \Phi_3(x)\} \vdash_{T^{iisi}} \Phi_1 \overset{typ}{\rightarrow} \Phi_3$

(b) $\forall x\{\Phi_1(x) \rightarrow \Phi_2(x)\},\ \Phi_3(iisi_{\Phi_2}),\ Unknown(\neg\Phi_3(iisi_{\Phi_1})) \vdash_{T^{iisi}} \Phi_3(iisi_{\Phi_1})$

Or, in $\overset{typ}{\rightarrow}$ notation, $\forall x\{\Phi_1(x) \rightarrow \Phi_2(x)\},\ \Phi_2 \overset{typ}{\rightarrow} \Phi_3,\ Unknown(\neg(\Phi_1 \overset{typ}{\rightarrow} \Phi_3)) \vdash_{T^{iisi}} \Phi_1 \overset{typ}{\rightarrow} \Phi_3$

Proof:

(a) Follows immediately using substitution and modus ponens.

(b) $\Phi_1(iisi_{\Phi_1})$ is an axiom by schema **A**. $\Phi_2(iisi_{\Phi_1})$ then follows from part (a) above. Thus by the theorem 3.1, $\Phi_3(iisi_{\Phi_1})$. $\square$

[9] In our more suggestive notation, $(Bird \overset{typ}{\rightarrow} Winged)$ follows from $(Bird \overset{typ}{\rightarrow} Flies)$, $(Flies \overset{typ}{\rightarrow} Winged)$, and $Unknown(\neg(Bird \overset{typ}{\rightarrow} Winged))$.

3.1.3 Some observations about the *iisi*-constant theory

Much like basic circumscription, the bare theory presented here cannot adjudicate between prioritized, competing, or interacting defaults. Some additional machinery would be required to give intuitive results when an ordering relation (be it subset-superset, chronological, etc.) exists between interacting defaults, but we see no principled reason to suspect that this cannot be done.

On the other hand, one feature (among others) that distinguishes the present theory from circumscription (when the predicate Ab is used to write defaults)[10] is the subtle difference in the way the two formalisms treat transitivity. In the *iisi* approach, defaults combine to reflect truly transitive reasoning: from "typically P's are Q's" and "typically Q's are R's" conclude "typically P's are R's". On the other hand, circumscription's Ab-default representation produces something like "typical P's which are typical Q's are R's" from the same initial defaults. More formally, consider the Ab-style defaults below:

$$\forall x\{P(x) \wedge \neg Ab_1(x) \;\rightarrow\; Q(x)\} \tag{3.1}$$

$$\forall x\{Q(x) \wedge \neg Ab_2(x) \;\rightarrow\; R(x)\} \tag{3.2}$$

$$\forall x\{P(x) \wedge \neg Ab_1(x) \wedge \neg Ab_2(x) \;\rightarrow\; R(x)\} \tag{3.3}$$

$$\forall x\{P(x) \wedge \neg Ab_3(x) \;\rightarrow\; R(x)\} \tag{3.4}$$

Defaults (3.1) and (3.2) state roughly that "typically P's are Q's" and "typically Q's are R's", respectively. Default (3.3) follows from (3.1) and (3.2), but (3.4) does not (where Ab_3 is an Ab-normality predicate intended to relate P and R). The difference between the two is that (3.3) reflects a more "cautious" sort of default composition than (3.4); cautious because it notes that defaults (3.1) and (3.2) are being composed by conjoining the Ab-normality predicates which appear in their antecedents. Default (3.4) is more akin to the *typ*-style transitivity discussed above. We will not specifically argue the pros and cons of (3.3) vs. (3.4); in a later section we further discuss abnormality predicates in circumscription.

We have now laid out the basic *iisi*-constant theory. In the next section we continue with an analysis of the theory, focusing on an apparently new kind of default, leading to a major and instructive shortcoming of the *iisi*-constant approach.

[10] We remind the reader of the standard use of Ab for *abnormal* [18]. In standard practice Ab predicates are indexed either by subscripts or by *aspect* constants. We use the former technique here for now, switching to aspect constants later when convenient.

4 Ranges and the Need for Denials: A Problem

The typical person has two eyes, two arms, and a mouth; is not an infant, and is not the President of the United States. Thus $iisi_{\text{Person}}$ is partially characterized by

$$Has\_Two\_Eyes(iisi_{\text{Person}}) \ \wedge \ \ldots \ \wedge \ \neg President\_of\_the\_US(iisi_{\text{Person}})$$

But problems can arise. Continuing with the above example, the typical person is singularly gendered—is male or female[11]—yet has no specified gender. There are too many male *and* too many female people to exclude either maleness or femaleness as a likely possibility of the typical person. Thus both of the defaults "typically people are male" and "typically people are female" are too restrictive and hence are inappropriate; i.e., "people are *not* typically male; and also *not* typically female" is also a piece of commonsense knowledge.[12] Indeed, the knowledge that people commonly come in either gender is such an everyday fact that we should view with suspicion any formalism that does not provide for its expression. In the gender default for people there is a *range* (i.e., the disjunction "male or female") which serves as a possible default conclusion, and this range cannot be restricted further.

4.1 Range defaults

Perhaps the most general form of what we call *range defaults*[13] is partly obscured in the example above by the fact that the class $Person$ is normally thought of as being exhausted by the two subclasses $Male$ and $Female$ (i.e., $Male \leftrightarrow \neg Female$). Consider a second example: typical wood has a color. It is either tan like pine or (dark) brown like walnut, though it's hard to pin it down any more than this. There is just too much tan and too much brown wood to exclude either from the range of a default about wood color. But clearly this does not exhaust the possibilities for wood; there is black, pink, red, green, and even purple wood, though none are prevalent enough to warrant recognition in the range of the wood-color default.[14]

In general, any number of non-exhaustive subclasses of class C can participate as disjuncts in the range of a default about C, which also asserts that any restriction of the disjunctive range is denied.

[11] Note that this is a default, not a universal fact. There are rare hermaphroditic or neuter persons who are neither (singularly) male nor (singularly) female.

[12] Note that this knowledge is used every day, and perhaps even more explicitly so in this era of political correctness. People often correct one another in particular over inappropriate use of the "masculine mystique" default: "typically people are male."

[13] This terminology was introduced in [23], although *irreducible disjunctive defaults* is more descriptive. For brevity we retain our original usage. To forestall a possible confusion we note that the well-known GCWA of Minker[26] also deals with default reasoning in the presence of disjunctive information. However, the ways disjunctions occur, and the kinds of inferences drawn, are quite different in GCWA and in range defaults.

[14] Notice that if we did include all wood colors, we would no longer have a default at all. In general, if the disjuncts in the range of a default about Ψ exhaust the category Ψ, then we no longer have a default.

In other words, a range default is an accepted default of the form "X's are typically Y's" where Y is a disjunction *and* for every shorter disjunction Z formed from the (disjunctive) components of Y, the (sub-)default "X's are typically Z's" is denied.

How can range defaults be formally represented? Let I be any non-empty finite set of (at least two) indices used to specify the disjuncts in a range and let $\bigvee_{i \in I} \Phi_i$ be the range of a default about Ψ: "typically things that satisfy Ψ also satisfy $\bigvee_{i \in I} \Phi_i$". Using our $\stackrel{typ}{\to}$ notation we denote the associated range default by[15]

$$\Psi \stackrel{typ}{\to} \boxed{\bigvee_{i \in I} \Phi_i} \tag{$\star$}$$

By $(\star)$ we intend two things. Firstly, that the plain disjunctive default principle

$$\Psi \stackrel{typ}{\to} \bigvee_{i \in I} \Phi_i \tag{4.1}$$

obtains and secondly that the range cannot be restricted any further. That is, for every non-empty proper subset J of I the potential default principle

$$\Psi \stackrel{typ}{\to} \bigvee_{i \in J} \Phi_i \tag{4.2}$$

is *denied*:[16]

$$\neg(\Psi \stackrel{typ}{\to} \bigvee_{i \in J} \Phi_i) \tag{4.3}$$

4.1.1 Range defaults and *iisi*-constants

Though useful in some respects, it turns out that the *iisi*-constant formalism is too weak represent range defaults. The intuitive reason is that the associated *iisi*-constant would need to have the disjunctive range property as well as fail to satisfy each and every disjunct. This produces a contradiction. The following theorem makes this formal.

Theorem 4.1 Let R be a range default expressed in terms of a *iisi*-constant, then R is inconsistent.

Proof: Straightforward. We illustrate the proof for the case of the gender range default. Here sentence (4.1) becomes

$$Person(x) \stackrel{typ}{\to} Male(x) \vee Female(x) \tag{4.4}$$

[15] We find the following iconic representation visually useful, although we will expand it immediately into the conjunction of statements (4.1) and (4.3) below.

[16] This notion of denial is reminiscent of, but distinct from, Pollock's [31] notion of an undercutting defeater. Undercutting qualifies the appropriate use of a default. Denial outright eliminates it as a default.

and (4.3) yields

$$\neg(Person(x) \overset{typ}{\rightarrow} Male(x)) \text{ and } \neg(Person(x) \overset{typ}{\rightarrow} Female(x)) \tag{4.5}$$

where $\{male\}$ and $\{female\}$ are the two non-empty proper subsets of $I = \{male, female\}$. Using an *iisi*-constant, sentence (4.4) becomes

$$Male(iisi_{\text{Person}}) \vee Female(iisi_{\text{Person}}) \tag{4.6}$$

which states that the typical person is either male or female, and (4.5) is

$$\neg Male(iisi_{\text{Person}}) \text{ and } \neg Female(iisi_{\text{Person}}) \tag{4.7}$$

which contradicts (4.6). □

This then captures in the *iisi*-constant formalism the oddity present in asserting that the typical person is singularly gendered, yet is not male and is not female. Theorem 4.1 states that in the *iisi*-constant formalism the oddity is present, and results in an inconsistency, regardless of the number of disjuncts in the range of the default.

To see further the cause of the contradiction notice that there is a strong intuitive pull toward two different readings of each of the statements in (4.7). Consider $\neg Male(iisi_{\text{Person}})$. One reading denies the masculine mystique, which we might denote by $\neg[Male(iisi_{\text{Person}})]$, and the other denies the maleness of the typical person, suggestively indicated by $[\neg Male]((iisi_{\text{Person}})$, itself a new mystique: the *non-masculine* mystique. However, formally these two readings are indistinguishable in any standard logic.

So the *iisi*-constant formalism has encountered a seemingly insurmountable difficulty: the inability to express range defaults. This suggests that fictitious *iisi*-entities may not be good choices for a formal treatment of default reasoning despite any psychological evidence that people use similar mental representations[8]. As previously mentioned, formal treatments of fictitious entities are tricky. The above is one more illustration of this. *Informal* commonsense reasoning allows the use of fictitious entities, and people have no trouble understanding that the associated contradictions of the sort shown above are not serious, because the entity is being used to represent a *population trend*. But the *formal* use of fictitious entities in an attempt to capture the distinction between the trend statements "people tend to be non-male" and "it is not the case that people tend to be male" ends up obliterating the distinction. The importance of an underlying trend statement will surface again, in a different way, below. Soon we will see that other formalisms found in the literature encounter related difficulties, albeit somewhat less severe, and we will develop a uniform simple treatment for these.

4.1.2 Why bother?

A natural question to ask is why bother with range defaults, or more precisely, with default denials at all? Why not just represent the default ($\star$) by sentence 4.1 and ignore marking the sentence (4.2) as inappropriate? After all, a sentence like (4.1) will give the right result for any instance of a Ψ-object. Knowing that a particular unpainted chair is made of wood and that typically wood is tan or brown should lead one to conclude (unless she knows to the contrary) that the chair is tan or brown. A reasoner need not go on to deny the individual defaults "typically wood is tan" and "typically wood is brown" in reaching that conclusion.

The answer is that default denials play a role in reasoning that is not accomplished solely with "ordinary" defaults. For instance there are cases in commonsense reasoning where it is not only important to reach the correct default conclusion, but also to have meta-knowledge about one's own defaults which itself can be reasoned with and about. This role is shown using a third illustration of a range default, that cardinals typically are red or russet (rust-colored).[17] Albino cardinals are exceptional. So it is indeed a commonsense disjunctive default. Moreover it is a range default since we do not accept (if we know much about cardinals!) that cardinals typically are red, nor that cardinals typically are russet. There are proportionately too many red and too many russet cardinals for either to be useful defaults in most domains.

Now, suppose you look out into the back yard of your house and notice that many red and no russet cardinals have gathered to eat. The plain disjunctive default that cardinals typically are red or russet does not prompt the conclusion that the collection of birds in your back yard is in any way unusual. But the event seems odd indeed, and you may have excellent reason to suspect so. Namely that it is *not* the case that cardinals typically are red. With no real knowledge about cardinal feeding or social behavior, it is appropriate to wonder at the oddity, and speculate that perhaps only red cardinals leave the nest to feed, or that only male cardinals like the trees in your yard. Such wondering and speculation can be prompted by the additional denial of the default that cardinals typically are red.

The above sort of knowledge that one may have about cardinals is precisely what a range default about cardinal color expresses, and this knowledge is crucial to the reasoning illustrated. Another illustration is to inform another reasoner to avoid a mistaken default; e.g., reminding someone that "people are not typically male." Thus, not only is the formal representation of range defaults of interest in a purely theoretical sense (Can a formalism represent them?), it also has pragmatic ramifications for commonsense reasoning formalisms.

The representation and use of range defaults clearly hinges on the representation and use of denied

[17] Typically males are red, females russet.

defaults: defaults that not only are not part of one's belief base but are explicitly believed to be false. This then is the topic we turn to next. However, we retain the larger picture of range defaults since they will play a role in how we go about solving the representational problems.

4.2 Denying defaults in other formalisms

The difficulty in representing range defaults is not peculiar to the *iisi*-constant approach. Let us return to the example of the gender range default, and consider how to represent it in three other widely studied formalisms, namely default logic (DL), nonmonotonic logic (NML), and circumscription. In each of the three formalisms we can write the plain disjunctive default principle. In NML the default is written as the axiom

$$\forall x\{Person(x) \wedge Consis(Male(x) \vee Female(x)) \ \rightarrow \ Male(x) \vee Female(x)\}. \tag{4.8}$$

In DL the default is written as the inference rule

$$\frac{Person(x) \ : \ Male(x) \vee Female(x)}{Male(x) \vee Female(x)} \tag{4.9}$$

where $\frac{P \ : \ Q}{R}$ has the approximate reading: If P is known, and if Q is consistent with all that is known, then R is inferred (known). In circumscription the default again is an axiom, as in NML, this time using an abnormality predicate, Ab:[18]

$$\forall x\{Person(x) \wedge \neg Ab_{Person;\{\mathrm{Male,Female}\}}(x) \ \rightarrow \ (Male(x) \vee Female(x)\} \tag{4.10}$$

In each formalism, the plain disjunctive default principle together with $Person(Tony)$, for instance, can be used to produce $Male(Tony) \vee Female(Tony)$, which is as expected. However, this does not get us home free. How can we represent the denials of the feminine and masculine mystiques? When we use the simple negation of the inappropriate (false) default (as we did above for range defaults using *iisi*-constants) problems arise in each formalism. We discuss some of these in what follows.

[18]Here we take up the practice of indexing Ab predicates with the names of the other predicates which appear in the default.

4.2.1 Denial in DL

A technical problem surfaces in DL, because here defaults are represented not as formulas but as rules of inference. Here the masculine mystique could be, and is in standard treatments, written as

$$\frac{Person(x) \ : \ Male(x)}{Male(x)} \tag{4.11}$$

But in DL (as in any standard logic) there is no recognized formal notion of the *negation* of a rule of inference.[19] Thus there appears to be no built-in way in DL to deny the masculine mystique using the standard representation of defaults. We will see later that a very simple change in default representation will allow DL to do this, however.

4.2.2 Denial in NML

Technical machinery is not the only thing getting in the way of default denial representations. Defaults are not merely declarative descriptions of facts about the commonsense world, such as "virtually all birds in these parts can fly." They also serve as *procedures* for basing defeasible inferences on those facts, such as the inference that "the bird over there can fly." It is duck soup to deny the former: "It is not the case that virtually all birds in these parts can fly." But denying a procedure is harder as we'll see in both NML and circumscription.

In NML, the standard representation of our ill-reputed masculine mystique default is

$$\forall x\{Person(x) \wedge Consis(Male(x)) \ \rightarrow \ Male(x)\} \tag{4.12}$$

Its negation then is

$$\neg\forall x\{Person(x) \wedge Consis(Male(x)) \ \rightarrow \ Male(x)\} \tag{4.13}$$

which is equivalent to

$$\exists x\{Person(x) \wedge Consis(Male(x)) \wedge \neg Male(x)\} \tag{4.14}$$

Intuitively (4.14) does not assert that the default itself is a bad one that should not be used so much as that there exist counterexample (at least one) to (4.12). But counterexamples should be expected now and then as simple evidence that (4.12) is a default after all, and not a universal fact. Thus a default and counterexamples to the default should be able to *coexist* rather than the counterexample negating (denying) the default.[20]

[19] Delgrande [5] makes this same point.

[20] This is addressed using a scoping mechanism in [10] and structural circumscription in [35]. However, neither scoping nor structural circumscription meets our needs: we want to avoid the default, not keep using it. In our own proposal below, we

This means that a denial is not the same as a counterexample: a default denial should go together with the *absence* of the default. Yet in traditional formalisms defaults are codified in terms of drawing conclusions about *individuals*: if such-and-such is the case, then conclude that a particular individual x has property P. Denying *this* leads directly to counterexamples (except in DL, where denials are not representable using the standard treatment at all). So we are in a quandary; we will analyze and attempt to resolve this in Section 6 below. In brief preview the issue seems to revolve around the difference between individual-oriented default representations (as in DL, NML, and circumscription) and population-oriented default descriptions. (For related comments see [21], p. 44, where a distinction due to Scriven is described.)

4.2.3 Denial in circumscription

Circumscription also has problems with denied defaults, much like those of NML. In the masculine mystique case, instead of the NML default axiom (4.12) we have

$$\forall x\{Person(x) \wedge \neg Ab_{Person;\{Male\}}(x) \;\rightarrow\; Male(x)\} \tag{4.15}$$

Here again the issue is whether the outright negation of this, namely

$$\exists x\{Person(x) \wedge \neg Ab_{Person;\{Male\}}(x) \;\wedge \neg Male(x)\} \tag{4.16}$$

really counts as denying the undesired default; the same arguments we gave earlier for NML apply.

Things only get worse when we try to reset the above in the original range (irreducible disjunctive) context. If a default has $\bigvee_{i\in I}\Phi_i(x)$ on its right hand side and $|I| = n$ then not only are $2^n - 2$ denials needed (just as in NML) but so are $2^n - 2$ additional abnormality predicates, one for each of the proper non-empty subsets of I.[21] (The proposed solution given in Section 6.3 may lend itself to a compaction of these exponentials associated with the sub-defaults of a range default.) More formally, for a default of the form

$$\forall x\{\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi_i\}_{i\in I}}(x) \;\rightarrow\; \bigvee_{i\in I}\Phi_i(x)\} \tag{4.17}$$

we require the negation of each of the more restrictive defaults of the form

$$\forall x\{\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi_i\}_{i\in J}}(x) \;\rightarrow\; \bigvee_{i\in J}\Phi_i(x)\} \tag{4.18}$$

can have the best of both worlds: denial of a default will avoid default, whereas a mere counterexample (properly scoped) leaves the default (both use and mention) in place for further inferences.

[21] Note that n can be rather large; consider "typical" colors of crayon, for instance.

where $\emptyset \neq J \subset I$. Each of these denials is a counterexample axiom; i.e.,

$$\exists x\{\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi_i\}_{i\in J}}(x) \wedge \neg \bigvee_{i\in J} \Phi_i(x)\} \tag{4.19}$$

As an example, the gender range default (where I is {Male, Female}), including the negation of both the masculine and feminine mystiques, looks like

$$\forall x\{Person(x) \wedge \neg Ab_{Person;\{\text{Male,Female}\}}(x) \;\rightarrow\; (Male(x) \vee Female(x)\} \tag{4.20}$$

$$\exists x\{Person(x) \wedge \neg Ab_{Person;\{\text{Male}\}}(x) \wedge \neg Male(x)\} \tag{4.21}$$

$$\exists x\{Person(x) \wedge \neg Ab_{Person;\{\text{Female}\}}(x) \wedge \neg Female(x)\} \tag{4.22}$$

Thus, in every model of (4.20)–(4.22) there are two counterexample persons a and b with $\neg Male(a)$ and $\neg Female(b)$. The existence of a is asserted by (4.21) and the existence of b by (4.22). But much as noted in [10] the latter two counterexample axioms serve to block the use not merely of the mystiques but also of the desired default that people typically are gendered![22] The following theorem makes this precise, even when $\mid I \mid = n > 2$.

Theorem 4.2 Let Ψ and subscripted Φ's be predicate letters, let I be a finite set of indices with at least 2 elements, let D be the plain disjunctive default

$$\forall x\{\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi_i\}_{i\in I}}(x) \;\rightarrow\; \bigvee_{i\in I} \Phi_i(x)\},$$

let S be the denial set (of sentences)

$$\{\exists x(\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi_i\}_{i\in J}}(x) \wedge \neg \bigvee_{i\in J} \Phi_i(x)) \mid \emptyset \neq J \subset I\},$$

and let a be a constant. Then $S + D + \Psi(a) \not\models_{circ} \bigvee_{i\in I} \Phi_i(a)$.

Proof: There is a minimal model (with respect to $Ab_{\Psi;\{\Phi_i\}_{i\in I}}$) of $S + D + \Psi(a)$ in which $\bigvee_{i\in I} \Phi_i(a)$ is not true, namely any model M with a single domain element. Specifically, $\Psi(a)$ must be true, and in order for S to be true in M (i) for all J, $\emptyset \neq J \subset I$, $Ab_{\Psi;\{\Phi_i\}_{i\in J}}(a)$ must be false and (ii) for each $i \in I$, $\Phi_i(a)$ must be false. Now for D to be true (since $\Psi(a)$ holds, and since $\bigvee_{i\in I} \Phi_i(a)$ fails because each $\Phi_i(a)$ fails) then $Ab_{\Psi;\{\Phi_i\}_{i\in I}}(a)$ must *hold*;[23] this requirement is easily satisfied by interpreting

[22] There are possible ways around this, of course [10, 35]. However, the ones we are aware of involve yet further complicating of the formalism. On the other hand, these remarks are intended as merely motivational; and our solution below avoids this entire issue by re-structuring the representation of default information in a very simple way.

[23] Note: $J \neq I$, hence this is a distinct Ab predicate from the ones in S, all of which fail in M.

$Ab_{\Psi;\{\Phi_i\}_{i\in I}}$ so that $Ab_{\Psi;\{\Phi_i\}_{i\in I}}(a)$ is true. These choices make M a model of the axioms $S+D+\Psi(a)$. Moreover, M is minimal since the extension of $Ab_{\Psi;\{\Phi_i\}_{i\in I}}$ cannot be made smaller without violating the truth of D. $\square$

Theorem 4.2 shows that in the presence of the denial set S, default reasoning is blocked. From $\Psi(a)$ and the default D one would expect the conclusion $\bigvee_{i\in I}\Phi_i(a)$, yet this is not forthcoming because $\neg Ab_{\Psi;\{\Phi_i\}_{i\in I}}(a)$ is not entailed. That is, we cannot show a to be normal even though we know nothing about a that would make it abnormal.

Another difficulty with this representation is that the meaning of the expression $\neg Ab_{\Psi;\{\Phi_i\}_{i\in J}}(x)$, found in each counterexample axiom, is unclear. In the absence of a believed default, what does $\neg Ab_{\Psi;\{\Phi_i\}_{i\in J}}(x)$ mean? Thus, what is it about a person and maleness that makes someone not abnormal (i.e., normal), as asserted in the above denial of the masculine mystique (4.21), if there is no such believed default asserting what is "normal maleness" for people, i.e., a default like (4.15), in the first place? An abnormality predicate is the standard circumscriptive mechanism for asserting what is normal (and abnormal); and the meaning of such assertions is standardly given in default axioms such as (4.15). The absence of a default (like 4.15), not to mention the presence of its denial (like 4.21), together doubly rob us of the essential characterization of the meaning of the default's abnormality predicate! We will return to this in Section 6.1.

Thus circumscription has the same problems with range defaults as does NML, along with some puzzling features regarding the semantics of the various abnormality predicates. In short, denial seems to create varied difficulties for most of the widely regarded default formalisms. While there are ways of handling some of these problems in isolation, we will propose an alternate approach that appears to handle all of the denial problems simultaneously and uniformly across all the formalisms we will consider in this paper.

4.3 Disjoined defaults

Reasoning about defaults is not limited to their denial. For example, consider remembering that there is a sport that Canadians typically love, and that it is either hockey or soccer, but being unsure of which. If Pierre is a Canadian we might reasonably conjecture that either Pierre loves hockey or Pierre loves soccer. Notice that this is very different from our earlier singularly-gendered (male or female) default: the Canadian example consists of two separate defaults at least one of which is taken to be true. In particular if Jeanette is also Canadian, then we conjecture Jeanette and Pierre both love the same sport. There is no implicit denial in this case; nevertheless Default Logic has no way to represent the disjunction of two defaults.

5 Related Work

The general issue of denying and otherwise reasoning *about* defaults has not gone unnoticed. Three existing approaches are those of Delgrande, Poole, and Schlechta, which we discuss below.

5.1 Delgrande's approach

Delgrande has developed default formalisms based on conditional logic in which, among other things, attention is paid to the representation of the denial of defaults (see [4] and [5]). In these logics—there are two of them, one propositional and one first-order—the conditional operator $\Rightarrow$ is added to an otherwise standard logical language in order to represent defaults, or what Delgrande calls statements of "normality." In the propositional version, called **NP**, a default is a statement of the form $\alpha \Rightarrow \beta$; intended to be read roughly as "if α then normally β." For example, "if it is raining then normally the grass is wet." In **N**, the first-order version, defaults are written in the form $\forall x(\alpha \Rightarrow \beta)$, intended to be read as "α's are normally β's." For instance, "ravens are normally black."

A possible world semantics is the basis for truth in both **NP** and **N**. Specifically $\alpha \Rightarrow \beta$ is true if β is true in each of the least exceptional worlds in which α is true, and $\forall x(\alpha(x) \Rightarrow \beta(x))$ is true in a world w iff $(\alpha \Rightarrow \beta)$ is true for each individual in the domain of individuals in w. In this formalism the denial of $\forall x(\alpha(x) \Rightarrow \beta(x))$ is simple and straightforward: $\neg\forall x(\alpha(x) \Rightarrow \beta(x))$ For instance, the denial of the male mystique is $\neg\forall x(Person(x) \Rightarrow Male(x))$.

Delgrande's primary aim in these logics is to provide a mechanism for reasoning *about* defaults, for drawing conclusions which themselves are defaults, rather than reasoning *with* them (e.g., for drawing conclusions about individuals). As examples, in the propositional case, the default $P \Rightarrow Q$ follows from $P \rightarrow Q$, and in the first-order case $\forall x(P(x) \Rightarrow R(x))$ follows from $\forall x(P(x) \Rightarrow Q(x))$ and $\forall x(Q(x) \rightarrow R(x))$.

Indeed, the first-order version **N** does not permit default reasoning about *individuals*. This is due to the fact that despite the suggestive notation, the default $\forall x(\alpha \Rightarrow \beta)$ is not to be construed as concerning ordinary quantification. Specifically, no mechanism is provided for substituting individuals for x, nor therefore for drawing default conclusions about an individual x from the default. Thus, though intuition may suggest that $Flies(tweety)$ ought to follow from $\forall x(Bird(x) \Rightarrow Flies(x))$ and $Bird(tweety)$, if Tweety is "normal", such is not the case in **N**. There simply is no inference rule nor axiom (schema) in the logic to help accomplish this. This prevents the problems seen in the previous section of the conflation of counterexamples with denials. Delgrande provides a separate mechanism for drawing default conclusions about individuals.[24]

One interesting feature of Delgrande's approach is the relationship which arises between typicality entailment ($\Rightarrow$) and strict entailment ($\rightarrow$): $\forall x(Px \Rightarrow Qx)$ follows from $\forall x(Px \rightarrow Qx)$. One could argue whether this is appropriate. Certainly $\Rightarrow$ and $\rightarrow$ bear some relation to one another, as Delgrande notes, though just what that relation ought to be may not be clear. On the one hand, it may seem reasonable that if *all P's are Q's* then also typically P's are Q's, as comes out in **N**. On the other hand it can be considered misleading to assert the typicality statement when the universal implication is believed. Asserting that "typically birds are animals," since " all birds are animals," is surely misleading as the former encourages the listener to assume that some birds are not animals. Grice's [12] maxim "be maximally informative" supports this view, as does the claim espoused in [10] which states that it is in the very nature of a default to have counterexamples. We refer to the use of a typicality statement when the corresponding universal implication is believed, as "understated universality."

Delgrande also points out that in some systems such as Default Logic, $\forall x(Rx \rightarrow Bx)$ and $\forall x(Rx \Rightarrow \neg Bx)$ can "peacefully coexist" even though this seems counterintuitive. If all R's are B's, then how can R's typically be non-B's? The DL treatment would allow the axiom $\forall x(Rx \rightarrow Bx)$ and the default rule $\frac{R\ :\ \neg B}{\neg B}$. Here the default rule will never be used since the default simply never applies, and so will not conflict with the axiom $\forall x(Rx \rightarrow Bx)$. Yet the intuitive reading of the rule *should* have it conflict with the axiom. The behavior of **N** differs: From $\forall x(Rx \rightarrow Bx)$ follows $\forall x(Rx \Rightarrow Bx)$, as stated above.

Nevertheless, viewed at a finer level, the "peaceful coexistence" problem may re-appear in Delgrande's approach in two related ways. First, within a single possible world both $\forall x R(x) \Rightarrow B(x)$ and $\forall x R(x) \rightarrow \neg B(x)$ can hold. For instance consider the model having two worlds $w_1 = \{R(a), R(b), \neg B(a), \neg B(b)\}$ and $w_2 = \{R(a), R(b), B(a), B(b)\}$ where Ew_1w_2[25]. Then $w_1 \models \forall x(R(x) \Rightarrow B(x)) \wedge \forall x R(x) \rightarrow \neg B(x)$; the default holds since $w_2 \models \forall x R(x) \rightarrow B(x)$ and w_2 is more normal than w_1. Delgrande does not consider this to be counterintuitive since $\Rightarrow$ refers not to the world at hand but rather to a more normal world. He further states that normalcy in his sense is not a statistical notion; however his formal treatment does ultimately depend on statistics or quantity in its requirement that in some accessible world *all* elements behave normally.

Delgrande gives as example a ten-fold increase in gravity rendering all birds flightless, where still one might say birds normally fly. But to us this appears to hinge on an intuition that such an episode is an irregularity in a much larger time span during most of which birds do fly. Thus again there appears to be a vestige of statistics in the underlying intuitions.

[24] See [5] for more on this. Our approach below, by way of contrast, accomplishes both *about* and *with* default reasoning using a single mechanism (which is a minor addition to traditional default formalisms.)

[25] Delgrande's notation for w_2 being accessible from w_1.

Second, suppose that the known facts or axioms state that there are exactly one thousand distinct entities $a_1 \ldots a_{1000}$ and that all of them satisfy R and exactly 999 of them (all but a_1) satisfy B. Then every world in every model must satisfy these facts as well. Now intuitively given this situation Rs are normally Bs, even though there is no world in which *all* Rs are Bs. So for Delgrande we do not get the default $\forall x R(x) \Rightarrow B(x)$: Our knowledge of the single counterexample $\neg B(a_1)$ effectively blocks the default. This is because Delgrande's semantics requires strict compliance in some world: 100% of Rs must be Bs.

Our example may seem unusual in that we have supposed all atomic facts known, not what default reasoning is usually concerned with. However, usual default reasoning aims to draw conclusions about individuals whereas Delgrande's treatment turns this around to draw conclusions about defaults. Thus if a reasoner knows about an entire population of individuals, it might very well reason about what defaults to adopt from that knowledge.

5.2 Poole's approach

Poole [32] bases his work on the premise that reasoning is not a matter of deduction but of theory formation. He views default reasoning as an attempt to use a set of hypotheses and a set of facts to explain observations. The hypotheses used in an explanation play the role of defaults.

Poole uses a first-order language together with a set $\mathcal{F}$ of facts, a set Δ of names of possible hypotheses and a set C of constraints. Names of possible hypotheses are predicates of the same arity as the hypotheses they name. For each name in Δ there is a fact in $\mathcal{F}$ which links the name to the hypothesis. If, for example, we want to have as default that birds normally fly, we would put $birdsfly(x)$ in Δ as the name of a hypothesis. $\mathcal{F}$ would then contain $\forall x\ birdsfly(x) \rightarrow (bird(x) \rightarrow fly(x))$. The constraints in C are formulas that state when a particular hypothesis is not applicable. For example, if we do not want the $birdsfly(x)$ hypothesis to apply to ostriches, we would put $\forall x\ ostrich(x) \rightarrow \neg birdsfly(x)$ in C. Poole defines a scenario of $(\mathcal{F}, \Delta, C)$ to be a set $D \cup \mathcal{F}$ where D is a set of ground instances of Δ and $D \cup C \cup \mathcal{F}$ is consistent. An *explanation* of a closed formula g is a scenario that implies g.

As an example, let

$$\mathcal{F} = \{bird(tweety), ostrich(fred), \forall x\ birdsfly(x) \rightarrow (bird(x) \rightarrow fly(x)), \forall x\ ostrich(x) \rightarrow bird(x)\}$$

$$\Delta = \{birdsfly(x)\}$$

$$C = \{\forall x \; ostrich(x) \rightarrow \neg birdsfly(x)\}$$

With this, we can "explain" $fly(tweety)$. The set D contains $birdsfly(tweety)$ which lets us derive $bird(tweety) \rightarrow fly(tweety)$ from $\mathcal{F}$, and from that we can get $fly(tweety)$. However we cannot find an explanation for $fly(fred)$ because $ostrich(fred)$ and the hypothesis $birdsfly(fred)$ are inconsistent with the constraint in C.

Unlike the approaches of Delgrande and Schlechta (below), Poole's treatment combines reasoning *with* defaults and reasoning *about* defaults. The latter is achieved through the constraint mechanism. It appears however to suffer from the peaceful coexistence problem: $n \in \Delta$ and $\neg n \in C$ together simply lead to the default named n not being used; no particular tension arises. This amounts to the odd situation in which $\forall x R(x) \rightarrow \neg B(x)$ can be a fact even though $n(x) \rightarrow (R(x) \rightarrow B(x))$ is also a fact: a default with *no* true instances.

Poole's system cannot reason about constraints such as that w is not a constraint. This is a limitation, in that one can imagine the need for a robot to be advised that, say, all ostriches except Alfred cannot fly, i.e., $birdsfly(x)$ is constrained by all ostriches except one. While we can write the constraint $(ostrich(x) \wedge x \neq alfred) \rightarrow \neg birdsfly(x)$ this does not entail the *absence* of the constraint $x = alfred \rightarrow \neg birdsfly(x)$, and there seems no way to express this absence.

An additional difficulty is the equivalence problem. Suppose for example, that we have default names n and m corresponding to the defaults $\forall x \; n(x) \rightarrow (bird(x) \rightarrow flies(x))$ and $\forall x \; m(x) \rightarrow (feathered(x) \rightarrow flies(x))$, together with the fact that $\forall x feathered(x) \leftrightarrow bird(x)$. If it is also known that $ostrich(tweety) \wedge \forall x \; ostrich(x) \rightarrow bird(x)$ and if $\forall x \; ostrich(x) \rightarrow \neg n(x)$ is a constraint then $n(tweety)$ is blocked but $m(tweety)$ is not and so we can still conclude $flies(tweety)$. The "equivalence" of n and m is not recognized.

5.3 Schlechta's approach

Schlechta [37] represents defaults in first-order logic with a generalized quantifier. An open default such as "Normally birds fly" is taken to mean "Most birds fly" and is represented as $\nabla x \; bird(x) : fly(x)$ which says that there is an "important" subset of the set of birds, all of whose members fly. We will mainly discuss the simpler case of normal open defaults: "Normally things don't fly", represented as $\nabla x \neg fly(x)$. This says that there is an important subset of the set of all objects, all of whose members don't fly.

The notion of important subsets is captured in the semantics by a system $\mathcal{N}(M)$ of important subsets of the domain. $\mathcal{N}(M)$ consists of subsets of M such that the intersection of any two of the subsets is not the empty set (unless the domain is empty). This condition rules out contradictory important subsets and therefore rules out situations where both "normally $\phi(x)$" and "normally $\neg\phi(x)$" are true. Also ruled out are empty important subsets when the domain is not empty. The notion of truth follows that of first-order logic with additional inductive steps including: $\nabla x\phi(x)$ holds in $\langle M, \mathcal{N}(M)\rangle$ provided there is a set A in $\mathcal{N}(M)$, for all members a of which $\phi(a)$ holds. This captures the idea that for something to be normally true there has to be an important subset of the domain for which that property holds. Since it is not possible for an important subset of the domain to satisfy $\neg\phi(x)$ while $\forall x\phi(x)$ holds, we cannot have "peaceful coexistence" between $\nabla x\neg\phi(x)$ and $\forall x\phi(x)$.

The language is a first-order language augmented with the quantifier ∇ and with a set of axiom schemata. The equivalence property is obtained from the axiom schema: $\nabla x\phi(x) \wedge \forall x(\phi(x) \rightarrow \psi(x)) \rightarrow \nabla x\psi(x)$. Inconsistencies in the defaults can be detected using the axiom $\nabla x\phi(x) \rightarrow \neg\nabla x\neg\phi(x)$. Analogous axioms are present in the open default case where another axiom of interest is $[\nabla x\phi(x) : \psi(x)] \wedge \forall x(\phi(x) \wedge \psi(x) \rightarrow \vartheta(x)). \rightarrow \nabla x\phi(x) : \vartheta(x)$. This enables us to infer from "normally birds fly" and "things that fly have wings" that "normally birds have wings." However we also have: $\forall x\phi(x) \rightarrow \nabla x\phi(x)$ which (as mentioned earlier in the discussion of Delgrande) we find unintuitive: if all birds are animals, then it seems misleading to say that typically birds are animals.

Schlechta's approach enables us to reason about (and deny) defaults and to resolve conflicts between defaults in a principled way. But as with Delgrande's first-order proposal, we cannot use Schlechta's scheme to conclude facts about individuals. There is no axiom or inference rule that lets us infer from $\nabla x\neg fly(x)$ that $\neg fly(fred)$. We applaud his introduction of sets, and think this can fruitfully be made more explicit (see below).

6 Formalizing Trends

6.1 Use and mention components of defaults

The representational difficulties illustrated in section 4 we see as hinging on a conflation of *use* (inferential, procedural) and *mention* (descriptive, declarative) components of default reasoning. To be sure, explicit denials of defaults, or inferences of a default from other defaults, have been considered (by Delgrande, Poole and Schlechta) apart from default use concerning individuals. But we think that even ordinary default use has an implicit default mention aspect that is better brought out into the open.

Research interest in the procedure (use) of drawing defeasible conclusions has been so high that the declarative (mention) component of default information has been given only limited attention.[26] The

[26] This is not to say that default information was not represented, nor that such representation had no declarative features, nor that typicality mentions were never used; but rather that little concern—Delgrande, Poole, and Schlechta are notable exceptions—had been shown for the detailed interpretation of various declarative mechanisms, as opposed to

latter (mention) is a bit hard to convey using the familiar terminology which is so infected with the inferential capacity of the former (use). We have already made a stab at it above in earlier sections, especially with regard to denials; however, the following may serve to reinforce the idea and make it more general. It may be a fact (or even just a belief) about the world that most birds can fly, quite independently of any disposition to draw defeasible conclusions from it. A separate piece of reasoning (whether via a rule or another belief) is required before the general belief about most birds can be applied to a particular bird. In DL this is done with rules; in NML and circumscription it is axioms that, together with yet further apparatus, produce defeasible conclusions. A state of affairs in which most birds can fly we will call a *trend*, namely a trend of the population of birds. Thus the existence of a population trend is logically prior to the inference that an individual member of that population obeys that trend (i.e. is typical or normal).

For concreteness (see the discussion near the end of Section 4.2.3): the circumscription default axiom $\Psi(x) \wedge \neg Ab_{\Psi;\{\Phi\}}(x) \rightarrow \Phi(x)$ both encodes what is normal *and* how to draw normality conclusions about any particular individual x. Denying this via a simple negation has the odd consequence of *affirming* $Ab_{\Psi;\{\Phi\}}(x)$, that is, of affirming there is a notion of normality (and hence of abnormality). The attempt to deny a notion of what is normal results in asserting it! We see this as due to the unfortunate mix of two mechanisms, normality-*mention* and normality-*use*, further amplified below, at work in a single formula.

The expression "default reasoning" is perhaps misleading with regard to this distinction. The declarative (mention) component is not a default, it is a (presumed) fact about a population. The default is the disposition to base defeasible reasoning on that declarative fact. However, it has become so customary to regard defaults as "encoding" typicality information that we will not try to buck tradition here. We will, however, suggest terminology for the two parts: typicality-mention (the trend) and default-use (the individual inference). It is the latter that holds all the nonmonotonicity.

We are accustomed to writing the world-facts of typicality-mention, e.g., that most birds fly, as part and parcel of the default-use mechanisms. As a result, it becomes problematic how to deny what is often not explicit or at least not separately represented: the typicality-mention itself.

6.2 Summing up the problem

Since defaults have tended to be formally represented in combined use-mention form, then they contain a confusion between a default being true as a mention of general typicality about a population, and sound

considerable concern for what default inferences were drawn from those representation. What we have argued above is that default information is also used in other ways (denials) where typicality mentions are crucial. Furthermore, as we saw in the previous section, even default conclusions can be adversely affected when denials in range defaults are not properly represented.

(or true) vis-a-vis its particular conclusions. The negation of such a representation (when possible at all) then mixes together both the unsoundness of the inferences, and the denial of the typicality-mention, even if (as in the case of range defaults) only the latter is wanted. As a result it is not clear whether we are getting what is needed when we try to deny a default, and our earlier discussion above suggests we are not.

The key, then, is to view a default D as having two complementary features. One is a declarative statement, $Trend$, as a fact about the commonsense world. The other is the individual-inference mechanism IIM by which $Trend$ is used to produce default conclusions. It is IIM that involves a plausibility assessment. In the usual formalisms these two are conflated. What we need for default-denial is to negate $Trend$, not IIM. But when IIM and $Trend$ are combined in a single monolithic representation, this is problematic.

IIM asserts, defeasibly, that default conclusions are true. $Trend$ asserts, indefeasibly, that the given typicality relation does obtain.[27] But that relation can obtain without all particular conclusions defeasibly inferred from it also obtaining.

To be more concrete, we may believe "birds typically fly," and this belief is distinct from (although related to) our subsequent beliefs that Tweety (probably) flies. We might find out that we are wrong about Tweety and yet retain our trust in the commonsense wisdom that birds typically fly. We may also conclude that, in this case, the bird in question (Tweety) is simply not typical.

Thus, from the fact that typically birds fly, it does not follow as a logical consequence that Tweety flies. Rather it follows as a defeasible conclusion under the (perhaps tacit) defeasible assumption that Tweety is typical. So, it may be that Tweety does not fly even though we have concluded the opposite, and without jeopardizing the truth of the fact that typically birds fly.

Should we want to deny the default itself, we assert that the typicality-mention is false: it is not the case the birds typically fly. We do not need to undo the default-use since that cannot function without the information of the typicality-mention.

6.3 A formal solution

It remains for us to formalize the above ideas and illustrate them in NML, DL, and circumscription. Our proposed solution then is to "divide" default representations into two parts, corresponding to their factual-mention ($Trend$) and inferred-conjectural-use (IIM) aspects.

We propose invoking a new predicate expression, roughly $Trend(P, Q)$, that has the intuitive meaning

[27] This does not mean that $Trend$ cannot later be found wrong and retracted; it simply means that $Trend$ is believed. Our treatment does not preclude the inference of a trend as a default; however we will not develop this here.

that "typically P's are Q's."[28] This plays the role of $Trend$ above: it is the claim that the commonsense world has a certain typicality feature. It is not in itself a means to infer particular default conclusions about particular entities. The latter is done by IIM, which takes different forms in different formalisms. We present here ways to represent IIM in DL, NML, and circumscription. The idea is that IIM will invoke $Trend$, roughly as follows:

IIM: If $Trend$, then draw a $Trend$-ish default conclusion when possible.

This contrasts with the traditional representation in which the antecedent "If $Trend$..." is missing.

Here we suppose typicality-mentions to be stated as ordinary axioms, e.g., $Trend(Bird, Flyer)$. In rough terms what we have in mind is along the following lines:[29]

$$Trend(Bird, Flyer) \rightarrow (Bird(x) \hookrightarrow Flyer(x))$$

The antecedent is our representation of the factual trend that birds do indeed typically fly, and the consequent $Bird(x) \hookrightarrow Flies(x)$ sanctions the defeasible (not necessarily factual) conclusion about an individual member x of the set of birds. Thus we combine both mention and use, but separated into sub-formulas. Moreover the mention (trend) sub-formula will in general also appear on its own as a separate belief and modus ponens will then be brought to bear when default use is desired.

Notice that our formulation has a superficial similarity to that of Poole:

$$Birdsfly(x) \rightarrow (Bird(x) \rightarrow Flies(x))$$

Nevertheless there is a major difference: $Birdsfly(x)$ is a statement about an individual satisfying a typicality condition, and is not a population trend at all. In fact we recall that in Poole's system there need be no trend at all; peaceful coexistence is permitted. However, our approach requires a trend to be in place before a default is used. This is not to say that trends should not be inferred from facts about individuals; we alluded to this in our example of one thousand known entities (section 5.1) and is a major topic for future work.

Our notation requires attention to some formal details which we develop now.

[28] Regarded as a relation between the classes P and Q.

[29] We use $\hookrightarrow$ as a generic notation for IIM (defeasible inference about an individual) whether in Default Logic, circumscription etc. We have abandoned $\stackrel{typ}{\rightarrow}$ since we used it earlier to capture traditional use-mention conflations. On occasion, the equivalent formulation $Trend(Bird, Flyer) \wedge Bird(x). \hookrightarrow Flyer(x)$ will be used.

6.3.1 Reification of properties

We seek a satisfactory means to express something along the lines of

$$Trend(R, B) \rightarrow (R(x) \hookrightarrow B(x))$$

where R and B are terms in the antecedent and yet $R(x)$ and $B(x)$ are formulas. This can be done in a number of ways, including a quotation device as in [27]:

$$Trend(\text{"}R\text{"}, \text{"}B\text{"}) \rightarrow (R(x) \hookrightarrow B(x))$$

Here we instead employ lowercase letters r and b to designate reified properties corresponding to the formulas $R(x)$ and $B(x)$: $Trend(r, b) \rightarrow (R(x) \hookrightarrow B(x))$. We now illustrate this device as a simple modification to traditional formalisms; then we will discuss a way to make this even more useful by treating the reified terms as sets.[30]

6.3.2 Formalizing denial in DL

We replace the inference rule

$$\frac{Person(x)\ :\ Male(x)}{Male(x)} \tag{6.1}$$

by the rule

$$\frac{[\,Trend(person, male) \wedge Person(x)\,]\ :\ Male(x)}{Male(x)} \tag{6.2}$$

Then as long as $Trend(person, male)$ is not believed, the rule cannot fire. We can even explicitly deny this typicality-mention by asserting $\neg Trend(person, male)$. Here $\hookrightarrow$ corresponds to the DL inference rule itself rather than to a formula.

6.3.3 Formalizing denial in NML

In NML we replace the axiom

$$Person(x) \wedge Consis(Male(x)) \rightarrow Male(x) \tag{6.3}$$

by

$$[\,Trend(person, male) \wedge Person(x) \wedge Consis(Male(x))\,] \rightarrow Male(x) \tag{6.4}$$

[30] Sets would appear to have a number of potential uses in commonsense reasoning; see [28, 29]

and argue as above for DL. Again, the default can be denied by asserting $\neg Trend(person, male)$, and without the conflation with counterexamples. Here the $\rightarrow$ corresponds to $\hookrightarrow$.

6.3.4 Formalizing denial in circumscription

Here we replace the axiom

$$Person(x) \wedge \neg Ab(x, i) \rightarrow Male(x) \tag{6.5}$$

where i is the relevant "aspect," by

$$[\, Trend(person, male) \wedge Person(x) \wedge \neg Ab(x, trend(person, male))\,] \rightarrow Male(x) \tag{6.6}$$

and again argue as above.[31] Again $\rightarrow$ corresponds to $\hookrightarrow$.

Note that $\neg Ab(x, i)$ is also a typicality statement, but about the individual x. That is, it says that x has the typical behavior for the population of people with respect to maleness, but it does not say what that typical population behavior is; $Trend(person, male)$ says the latter.[32]

6.3.5 Using sets

We now return to the idea of reifying population trends as sets. This has the advantage that now we can give an arithmetic characterization of population trends, namely $Trend(r, b)$, will be true if "most" of the set r is contained in the set b. This can be formalized by a parameter λ that measures the proportion of rs that are bs. We say that $Trend(r, b)$ is true iff

$$\lambda < \frac{|r \cap b|}{|r|} < 1$$

Here the first inequality expresses that most of r is contained in b, and the second inequality addresses the understated universality issue: we take a typicality statement to entail that there are exceptions. [33]

The terms r in the typicality formulas are a notational simplification for $\{x \mid R(x)\}$. To relate these

[31] We find it happily suggestive here to interpret McCarthy's "aspects" as simply the population trends themselves ($Trend(p, q)$ reified via a function symbol as $trend(p, q)$).

[32] Vladimir Lifschitz has pointed out to us that we can bend the Ab predicate to our needs, eliminating its individual-dependence, by writing $\forall x Ab(x, a)$ for some aspect a, to replace our $\neg Trend(P, Q)$. Then $Trend(P, Q)$ corresponds to $(\exists x)\neg Ab(x)$; but this has an odd feel to it as a way to assert a population trend: one "normal" entity is surely not enough for there to be a valid notion of normality, let alone *no* such entities as is suggested by use of $\forall x Ab(x, a)$—the peaceful coexistence issue again, where a trend is represented and yet applies to nothing at all.

Thus $Ab(x, a)$ seems to have a presupposition, that the indicated aspect encodes a genuine population trend (which however is not followed by x). Negating $Ab(x, a)$ then is confusing for the same reasons as we have dwelled in above: which of the two is being denied? In our treatment these are separated.

[33] We put aside the issue of how much of a set constitutes "most" of it; this can vary greatly depending on context. Presumably $\lambda > 0.5$; our treatment will *not* depend on the specific numeric proportion λ. In future work we intend to explore connections with other statistical foundations [1]. The second inequality can be left out with no harm done to our treatment; it simply serves to avoid understated universality.

terms to formulas $R(x)$, the language needs to be augmented:

1. We introduce a binary predicate $\in$ for set membership.

2. For each formula $\phi(x)$ in the language, we add a term

$$\{x \mid \phi(x)\} \tag{6.7}$$

   where x is one of the free variables of ϕ, representing the set of individuals satisfying the formula.

3. We employ a "comprehension" axiom schema:

$$\forall y(\phi(y) \leftrightarrow y \in \{x \mid \phi(x)\}) \tag{6.8}$$

   where $\phi(y)$ is a formula in the language: this relates the formula to the set term introduced above.[34]

4. For each typicality statement $Trend(\{x \mid \phi(x)\}, \{x \mid \psi(x)\})$ (which is written more simply as $Trend(\phi, \psi)$) we add the axiom:

$$Trend(\{x \mid \phi(x)\}, \{x \mid \psi(x)\}) \rightarrow \forall y(y \in \{x \mid \phi(x)\} \hookrightarrow y \in \{x \mid \psi(x)\}) \tag{6.9}$$

5. We also add the axiom schema

$$\{x \mid \phi(x)\} = \{x \mid \psi(x)\} \leftrightarrow (\forall z\ z \in \{x \mid \phi(x)\} \leftrightarrow z \in \{x \mid \psi(x)\}) \tag{6.10}$$

   for set equality.

As an illustration, suppose we are given the trend axiom:

$$Trend(bird, fly)$$

Then from 6.9 it follows that

$$\forall x(x \in \{x \mid Bird(x)\} \hookrightarrow x \in \{x \mid Fly(x)\})$$

[34] Introducing $\in$ into the language, together with schema (6.8), brings with it the possibility of paradoxes. There are many proposals for dealing with this, e.g. [27, 11, 2]. Here we opt for a very conservative solution to the problem: we suppose an original language L that does not contain the predicate symbols $\in$ or $Trend$. We then extend this with set constants $\{x \mid \phi(x)\}$ for each formula $\phi(x)$ in L, as well as with the two symbols $\in$ and $Trend$. We employ the comprehension schema 6.8 only for formulas ϕ in the original language L. This amounts to a second-order logic, a very weak set theory. Stronger set theories are also available, of course.

If tweety is a constant in the language, we then derive

$$tweety \in birds \hookrightarrow tweety \in fly$$

We can now solve the equivalence problem (which arose in Poole's system) in our augmented language. If we know that typically birds fly, $Trend(bird, fly)$, and we also know that birds are precisely the feathered vertebrates, $\forall x Bird(x) \leftrightarrow Fv(x)$, we want to conclude that typically feathered vertebrates fly, $Trend(fv, fly)$.

We have

$$Trend(bird, fly)$$

and

$$\forall x(Bird(x) \leftrightarrow Fv(x))$$

The first of these is simply

$$Trend(\{x \mid Bird(x)\}, \{x \mid Fly(x)\})$$

The comprehension schema yields the following two formulas:

$$\forall y(Bird(y) \leftrightarrow y \in \{x \mid Bird(x)\})$$

$$\forall y(Fv(y) \leftrightarrow y \in \{x \mid Fv(x)\})$$

Substitution then provides:

$$\forall y(y \in \{x \mid Bird(x)\} \leftrightarrow y \in \{x \mid Fv(x)\})$$

We then get the set equality:

$$\{x \mid Bird(x)\} = \{x \mid Fv(x)\}$$

Another substitution gives:

$$Trend(\{x \mid Fv(x)\}, \{x \mid Fly(x)\})$$

which is the same as:

$$Trend(fv, fly)$$

As a final illustration recall an example mentioned in our discussion of Delgrande: $\forall x(P(x) \Rightarrow R(x))$ follows from $\forall x(P(x) \Rightarrow Q(x))$ and $\forall x(Q(x) \rightarrow R(x))$. If it turns out that $\forall x(P(x) \rightarrow R(x))$ then, as we have already observed, we question the appropriateness of asserting that Ps normally are Rs:

the expressions normality and typicality carry a sense that most *but not all* members of $\{x \mid P(x)\}$ are members of $\{x \mid R(x)\}$. In our treatment because of the second inequality in $\lambda < \frac{|r \cap b|}{|r|} < 1$ from our characterization of $Trend(r, b)$, this sense is enforced. As a consequence we can prove the following strengthened version of Delgrande's result: suppose $Trend(p, q)$ and $\forall x(Q(x) \rightarrow R(x))$. Then $\lambda < \frac{|p \cap q|}{|p|} < 1$ by the statistical characterization of trends; it follows that $\lambda < \frac{|p \cap r|}{|p|}$ by elementary finite set theory. If in addition $\exists x(P(x) \wedge \neg R(x))$ then the second inequality follows as well: $\frac{|p \cap r|}{|p|} < 1$.

6.3.6 Exponentials avoided

If a powerful enough language is available (e.g., set theory as above), the exponentially-many to-be-denied sub-defaults (typicality-mentions) associated with a range default can be described compactly in generic form by means of quantifying over subsets of the n disjuncts in a range default. A similarly compact treatment can be given for desired default-use mechanisms. Thus all default-uses in, say, circumscription could be written at once in the form:

$$\forall u, v\ [Trend(u, v) \wedge x \in u \wedge \neg Ab(x, trend(u, v)). \rightarrow x \in v]$$

7 Discussion and Future Work

We have examined default reasoning from a number of perspectives, beginning with that of range defaults and *iisi*-constants. While such constants did not in the end prove workable, their study helped clarify the nature of the problem of denying default principles, a problem which remains in traditional formal approaches as well. Others who have considered default denials and other meta-default notions have proposed formalisms differing markedly from the traditional ones; moreover, those new proposals are themselves not without curious features that may not be appropriate. Our own proposal stays within traditional formal methods, and simply conjoins a new predicate to them. The new predicate plays a meta-role, which we have characterized as a population trend, while the remaining (traditional) portion now can be seen more clearly in its pure default-use role, eliminating previous conflation.

If a population-based notion of typicality is to be taken seriously, then a deeper connection between typicality-mentions and actual population data is called for. We noted earlier that Delgrande's approach does not offer this; our own "trend" approach as captured in the *Trend* predicate applied to sets does appear to lend itself to such an attempt. More work is needed to see how far this can be carried.

Recent work by Bacchus et al [1] provides a very satisfying tie between numerical (statistical or probabilistic) approaches to nonmonotonic reasoning, and qualitative approaches such as are found in the various logic-based formalisms. Our own efforts reported here can be seen as a different step in this direction: we have highlighted default-mention as quintessentially a matter of (a belief concerning) a population trend. We believe it essential to carry forward these ideas fully if nonmonotonic reasoning is to enter into the ranks of genuinely practical techniques in the AI arsenal. A reasoning agent that cannot reason (or be usefully advised) that one class under consideration is much larger than another, or that two classes are sufficiently similar in characteristics that they can be treated alike, or that so few members of one class have a given default property that the default should be declared inappropriate, is a very poor commonsense agent indeed.

The chief advantages we see for our approach are (i) that it allows default denials (and other metareasoning) by almost trivial adjustments to existing approaches, (ii) it appears to have the beginnings of a mechanism for a more proper connection with population data, and (iii) it clarifies what had been a monolithic conflation of two conceptually distinct notions.

Conflations are perhaps most notorious in the case of the *iisi*-constants we analyzed early in the paper, but appear as well in more traditional formalisms. And the apparent commonplace occurrence of range defaults makes it essential to tease the conflations apart.

We conclude by re-examining the conflation that occurs in the *iisi*-theory, which is parallel to but a bit different from that of use-mention. $\Phi(iisi_\Psi)$ masquerades as a typicality-mention (the population trend that Ψs are typically Φs). And yet prepending a negation sign conflates the denial of such a trend with the assertion of the opposite trend (Ψs are typically non-Φs). The denial would simply assert that it is not the case that Ψs are typically Φs. What is needed to achieve this is a syntactic marker that has a unique role as a trend-assertion. The extra baggage of *iisi*-constants gets in its own way and causes the negation to assert more than is wanted: it asserts that the constant has the negated property, because for an individual, failing to have a property is the same as to have its negation (at least in two-valued logics). Replacing the *iisi*-treatment with $Trend(\psi, \phi)$ achieves precisely what is wanted, turning attention to the population rather than a (fictitious) individual.

The same observations apply to a disjunctive trend (range default): instead of asserting that typically people are either male or female, the obvious disjoined *iisi*-formula asserts that the either the typical person is a male (and hence most people are males), or the typical person is a female (and most people are females). An *iisi*-constant does its job too well, forcing defaults when none is wanted. This closing example serves to remind that, despite the title of this paper, it is not denials alone that are problematic, and that appear well-served by the population-based approach we advocate.

As mentioned earlier, it is of interest that humans apparently do at times reason with something very like *iisi*-constants, yet do not seem to make the associated conflation errors. We suspect that the ability to judiciously refrain from applying a given mode of thought in certain contexts is a large part of commonsense reasoning, and one in need of much more exploration [13, 19, 30].

References

[1] F. Bacchus, A. Grove, J. Halpern, and D. Koller. Statistical foundations for default reasoning. In *IJCAI*, 1993.

[2] J. Barwise and J. Etchemendy. *The Liar: An essay on truth and circularity.* Oxford University Press, 1987.

[3] R. Brachman. I lied about the trees or, defaults and definitions in knowledge representaion. *AI Magazine*, 6(3):80–93, 1985.

[4] J. P. Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33(1):105–130, 1987.

[5] J. P. Delgrande. An approach to default reasoning based on first-order conditional logic: Revised report. *Artificial Intelligence*, 36(1):63–90, 1988.

[6] J. Elgot-Drapkin. *Step-logic: Reasoning Situated in Time.* PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 1988.

[7] J. Elgot-Drapkin and D. Perlis. Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2(1):75–98, 1990.

[8] H. C. Ellis and R. R. Hunt. *Fundamentals of Cognitive Psychology, 5th edition.* Wm. C. Brown Communications, 1993.

[9] D. Etherington. *Reasoning with Incomplete Information.* Research Notes in Artificial Intelligence. Morgan Kaufmann, Los Altos, CA, 1988.

[10] D. Etherington, S. Kraus, and D. Perlis. Nonmonotonicity and the scope of reasoning. *Artificial Intelligence*, 52:221–261, 1991.

[11] S. Feferman. Toward useful type-free theories–I. *Journal of Symbolic Logic*, 49:75–111, 1984.

[12] H. P. Grice. *Logic and Conversation.* William James Lectures. Harvard University, 1967.

[13] R. V. Guha. *Contexts: A formalization and some applications.* PhD thesis, Stanford University, Palo Alto, CA, 1991.

[14] G. Hirst. Existence assumptions in knowledge representation. *Artificial Intelligence*, 49:199–242, 1991.

[15] H. Kyburg. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, CT, 1961.

[16] J. McCarthy. First order theories of individual concepts and propositions. *Machine Intelligence*, 9:129–147, 1979.

[17] J. McCarthy. Circumscription: A form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39, 1980.

[18] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28(1):89–116, 1986.

[19] J. McCarthy. Notes on formalizing context. In *Second Workshop on Formal Common Sense Reasoning*, 1993.

[20] D. McDermott. Non-monotonic logic II. *Journal of the ACM*, 29:33–57, 1982.

[21] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13(1,2):41–72, 1980.

[22] D. L. Medin and B. H. Ross. *Cognitive Psychology.* Harcourt Brace Jovanovich, 1992.

[23] M. Miller and D. Perlis. Typicality constants and range defaults: Some pros and cons of a cognitive model of default reasoning. In *Proceedings of the 1991 SIGART International Symposium on Methodologies for Intelligent Systems*, 1991.

[24] M. Miller and D. Perlis. What experts deny, novices must understand. In *3rd International Workshop on Human and Machine Cognition*, Seaside, Florida, 1993.

[25] M. Miller and D. Perlis. Toward automated expert reasoning and expert-novice communication. In K. Ford, P. Feltovich, and R. Hoffman, editors, *Human and Machine Expertise in Context*. to appear.

[26] J. Minker. On indefinite databases and the closed world assumption. In *Proceedings of 6th Conference on Automated Deduction*, pages 292–308, New York, 1982.

[27] D. Perlis. Languages with self reference I: Foundations. *Artificial Intelligence*, 25:301–322, 1985.

[28] D. Perlis. Circumscribing with sets. *Artificial Intelligence*, 31:201–211, 1987.

[29] D. Perlis. Commonsense set theory. In P. Maes and D. Nardi, editors, *Meta-Level Architectures and Reflection*. North Holland, 1988.

[30] D. Perlis. Context: Theory and practice. *Fundamenta Informaticae*, 23, 1995. A special issue of eight papers by various authors, edited and with an introduction by D. Perlis.

[31] John Pollock. How to reason defeasibly. *Artificial Intelligence*, 57(1):1–42, 1992.

[32] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[33] W. Rapaport. How to make the world fit our language: An essay in Meinongian semantics. *Grazer Philosophische Studien*, 14:1–21, 1981.

[34] W. Rapaport. Logical foundations for belief representation. *Cognitive Science*, 10:371–422, 1986.

[35] P. K. Rathmann, M. Winslett, and M. Manasse. Circumscription with homomorphisms: Solving the equality and counterexample problems. *Journal of the ACM*, 41(5):819–873, 1994.

[36] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[37] K. Schlechta. Defaults as generalized quantifiers. *Journal of Logic and Computation*, 5(4):473–494, 1995.

[38] R. L. Solso. *Cognitive Psychology, 3rd edition*. Allyn and Bacon, 1991.