For all typical

Natasha Alechina

FWI, University of Amsterdam
Pl. Muidergracht 24, 1018 TV Amsterdam
The Netherlands
e-mail natasha@fwi.uva.nl

Abstract. In the paper a binary quantifier $\Pi x(\varphi, \psi)$ is studied. It can be read as 'typically, φ's are ψ's'. After investigating some formal properties of this quantifier, we compare the resulting logic with predicate conditional logic and discuss possible applications of binary quantifiers in formalizing defeasible reasoning.

1 Introduction

The present paper contains one more suggestion concerning a formal way to treat statements like 'φ's are (normally, typically) ψ's': common sense generalizations which can be used as premises in defeasible reasoning. The common feature of those statements is that they can be true while some exceptions (φ's which are not ψ's) are present. In the paper they are formalized with the use of a binary generalized quantifier Π, with $\Pi x(\varphi(x), \psi(x))$ informally meaning 'for all typical φ's ψ is true'. A related work in this respect is, first of all, Badaloni and Zanardo (1990), (1991). Another connection (which was already discussed by Badaloni and Zanardo) is the way such statements are formalized in the theory of circumscription: if $\varphi(x)$ and x *is not abnormal*, then $\psi(x)$. In section 4 possible applications of the logic with generalized quantifiers to defeasible reasoning are sketched. In section 3 we compare two possible readings of commonsense generalizations: 'normally, φ implies ψ' (cf. Morreau (1992)) and 'all typical φ's are ψ's'.

2 Language and models

Consider a first-order language $L_{\forall\Pi}$ with a binary generalized quantifier Π. A well-formed formula is defined as usual; if φ and ψ are well-formed formulas, so is $\Pi x(\varphi, \psi)$.

A model M for $L_{\forall\Pi}$ is a triple $< D, T, I >$, where D is a domain, I an interpretation function, and T is a function which takes a subset S of D and returns a set $S_T \subseteq D$ (the set of 'typical S's'). T obeys only one constraint, namely

$$S_T \subseteq S.$$

The truth definition for the generalized quantifier reads as follows:

$$M \models^\alpha \Pi x(\varphi, \psi) \Leftrightarrow \{d : M \models^\alpha \varphi[x/d]\}_T \subseteq \{d : M \models^\alpha \psi[x/d]\}$$

The notions of truth, validity and semantic consequence are standard. The following logical system corresponds to the described semantics.

Definition 1. The logic *Bin* is the least set of formulas closed under first-order derivability and

Reflexivity $\vdash \Pi x(\varphi, \varphi)$;
Distributivity $\vdash \Pi x(\varphi, \psi_1) \wedge \Pi x(\varphi, \psi_2) \rightarrow \Pi x(\varphi, \psi_1 \wedge \psi_2)$
Monotonicity $\vdash \forall x(\varphi(x) \rightarrow \psi(x)) \wedge \Pi x(\chi, \varphi) \rightarrow \Pi x(\chi, \psi)$;
Exchange $\vdash \forall y \Pi x(\varphi, \psi) \rightarrow \Pi x(\varphi, \forall y \psi)$, given that y is not free in φ;
Extensionality $\vdash \forall x(\varphi \leftrightarrow \psi) \rightarrow (\Pi x(\varphi, \chi) \leftrightarrow \Pi x(\psi, \chi))$;
Alphabetic variants $\vdash \Pi x(\varphi(x), \psi(x)) \rightarrow \Pi y(\varphi(y), \psi(y))$, given that y is free for x in $\Pi x(\varphi(x), \psi(x))$.

The notion of a proof is standard. Since in both logics the only inference rules are modus ponens and generalization, the deduction theorem has the same conditions (and proof) as in classical predicate logic.

It is easy to see that *Bin* is sound with respect to the semantics described above. A different question is whether the axioms are intuitively appealing given the intended interpretation of Π. Especially distributivity, exchange and extensionality can cause objections.

The first two axioms can be criticized as follows. There are reasonable interpretations of defaults under which even the closure for finite intersections in the consequent (distributivity) does not hold. For example, when $\Pi x(\varphi, \psi)$ is interpreted as 'in more than 50 per cent of cases φ's are ψ's'. Exchange axiom is even more questionable, since it presupposes the closure for arbitrary intersections. It fails under such strong interpretation of $\Pi x(\varphi, \psi)$ as $P(\{x : \psi(x)\}|\{x : \varphi(x)\}) = 1$. The following argument showing that the intersection property is counterintuitive, is given in Badaloni and Zanardo (1990). For every *specific* location y (for example, Rome), the generic 'Generally, birds live outside y' seems valid, that is, (after quantifying over locations)

$$\forall y \Pi x(Bird(x), Lives.outside(y, x))$$

can be accepted. But $\Pi x(Bird(x), \forall y Lives.outside(y, x))$ is obviously false.

However, there is an interpretation, namely the one underlying the semantics of *Bin*, which makes the axioms valid. If *there are* typical objects, then the intersection property must hold. The arguments above show that such 'absolute typicality' might be too strong an assumption. My suggestion would be to adopt the strategy used in the theory of circumscription and introduce 'typicality relative to an aspect'. Technically speaking, instead of one function T introduce a function for every aspect, $T_1, T_2, \ldots$, and corresponding quantifiers $\Pi^1, \Pi^2, \ldots$. The example above becomes $\Pi^1 x(Bird(x), Lives.outside(Rome, x))$, $\Pi^2 x(B(x), Lives.outside(Paris, x)), \ldots$, et cetera. The undesirable conclusion does not follow any more. To keep things simple, only one quantifier is considered in the sequel; but the analysis below will not change if the logic contained countably many quantifiers.

The extensionality property is also counterintuitive. If by chance the sets of A's and B's happen to contain precisely the same elements, it does not mean that typical A's are typical B's, and vice versa. To get rid of the extensionality, T can be defined syntactically, on formulas instead of sets; this is done in Alechina (1993) for a slightly different set up. But this also makes things too complicated, and for the present purposes we keep extensionality as well.

Theorem 2. *Bin is complete.*

Proof. The following principles, which are derivable in Bin, will be used in the proof:

Iteration $\vdash \Pi z(\varphi(z), \forall y_1 \ldots \forall y_n(\Pi x(\varphi(x), \psi(x)) \rightarrow \psi(z)))$, given that $y_1, \ldots, y_n$ are not free in φ (a derivation can be found in Alechina (1993)),

Monotonicity rule $\Sigma \vdash \varphi(x) \rightarrow \psi(x) \Rightarrow \Sigma \vdash \Pi x(\chi, \varphi) \rightarrow \Pi x(\chi, \psi)$, where x is not free in Σ.

We are going to show that every consistent in the sense of Bin set of formulas Δ has a model. Δ is extended to a maximally consistent set of formulas in a usual way, adding a witness for every existential quantifier. $\Sigma_0 = \Delta$, and the set of formulas which is obtained on the nth step will be denoted as Σ_n. If the $n+1$st formula is of the form $\neg\Pi x(\varphi, \psi)$ and it is consistent with Σ_n, add a new variable x' such that

(1) $\neg\psi[x/x']$;
(2) $\{\forall y_1 \ldots \forall y_k(\Pi x(\varphi(x), \chi(x)) \rightarrow \chi(x'))$:
for all formulas χ in the old alphabet, where $y_1, \ldots, y_k$ are the free variables of χ not occurring in $\varphi\}$

Assume that the above algorithm gives rise to inconsistencies. Then there is a finite subset of the formulas added on the nth step which is inconsistent with Σ_n, that is

$$\Sigma_n \vdash \bigwedge_i \forall y_{i1} \ldots \forall y_{ik}(\Pi x(\varphi(x), \chi_i(x)) \rightarrow \chi_i(x')) \rightarrow \psi(x')$$

and, by monotonicity,

$$\Sigma_n \vdash \Pi z(\varphi(z), \bigwedge_i \forall y_{i1} \ldots \forall y_{ik}(\Pi x(\varphi(x), \chi_i(x)) \rightarrow \chi_i(z))) \rightarrow \Pi z(\varphi(z), \psi(z))$$

and, by alphabetic variants,

$$\Sigma_n \vdash \Pi z(\varphi(z), \bigwedge_i \forall y_{i1} \ldots \forall y_{ik}(\Pi x(\varphi(x), \chi_i(x)) \rightarrow \chi_i(z))) \rightarrow \Pi x(\varphi(x), \psi(x)).$$

Since by iteration

$$\Sigma_n \vdash \Pi z(\varphi(z), \forall y_{i1} \ldots y_{ik}(\Pi x(\varphi(x), \chi_i(x)) \rightarrow \chi_i(z)))$$

for every χ_i, by Distributivity

$$\Sigma_n \vdash \Pi z(\varphi(z), \bigwedge_i \forall y_{i1} \dots y_{ik}(\Pi x(\varphi(x), \chi_i(x)) \to \chi_i(z))),$$

therefore $\Sigma_n \vdash \Pi x(\varphi(x), \psi(x))$. A contradiction.

The model is based on the set $\Sigma = \cup_{n<\omega} \Sigma_n$. As usual, $, d_1, \dots, d_n >\in I(P) \Leftrightarrow P(d_1, \dots, d_n) \in \Sigma$. T is defined as follows:

$$u \in \{d : \varphi(x/d) \in \Sigma\}_T \;\Leftrightarrow\; \forall \chi(\Pi x(\varphi(x), \chi(x)) \in \Sigma \Rightarrow \chi(u) \in \Sigma).$$

The constraint on T: $\{d : \varphi(x/d) \in \Sigma\}_T \subseteq \{d : \varphi(x/d) \in \Sigma\}$ holds because of the Reflexivity axiom.

T is a function, i.e. if $\{d : \varphi(x/d) \in \Sigma\} = \{d : \psi(x/d) \in \Sigma\}$, then $\{d : \varphi(x/d) \in \Sigma\}_T = \{d : \psi(x/d) \in \Sigma\}_T$: this follows from Extensionality.

The construction of Σ guarantees that for every formula ψ, if $\Pi x(\varphi(x), \psi(x)) \notin \Sigma$, then there exists $u \in \{d : \varphi(x/d) \in \Sigma\}_T$ such that $\neg\psi(u) \in \Sigma$. Take u to be the witness x' for $\neg \Pi x(\varphi(x), \psi(x))$. Assume that x' is not in $\{d : \varphi(x/d) \in \Sigma\}_T$, that is, there is a formula χ containing some variables which were introduced after x', such that $\Pi x(\varphi(x), \chi(x)) \in \Sigma$ and $\neg\chi(x') \in \Sigma$. Then $\exists z_1 \dots z_n(\Pi(\varphi(x), \chi(x)) \wedge \neg\chi(x')) \in \Sigma$, a contradiction. Now it is easy to prove that for any formula θ, $\theta \in \Sigma \Leftrightarrow \Sigma \models \theta$. $\square$

3 Binary quantifiers and conditionals

In this section we will study the connections between the representation of defaults considered above and their representation in predicate conditional logic (cf. Delgrande (1991), Morreau (1992)).

Let $L_{\forall >}$ denote the language of first-order conditional logic (the language of the first-order predicate logic plus binary modal operator $>$). M is *a model* for $L_{\forall >}$ if $M =< D, W, S, I >$, where D is a non-empty universe (the same for all possible worlds), W is a non-empty set of possible worlds, $S : W \times \mathcal{P}(W) \to \mathcal{P}(W)$ is a selection function (given a world and a proposition, S provides a set of worlds which are *normal* with respect to this proposition from the point of view of this world), and I is an interpretation function. The truth definition for conditionals is as follows:

$$M, w \models^\alpha \varphi > \psi \Leftrightarrow S(w, [\varphi]_{M,\alpha}) \subseteq [\psi]_{M,\alpha},$$

where $[\varphi]_{M,\alpha} = \{w' : M, w' \models^\alpha \varphi\}$. The selection function satisfies the following constraint:

ID $S(w, [\varphi]_{M,\alpha}) \subseteq [\varphi]_{M,\alpha}$

Definition 3. *Cond* is the least set of formulas closed under first-order derivability and the axioms and rules below:

CI $\varphi > \varphi$;

CC $(\varphi > \psi_1) \wedge (\varphi > \psi_2) \rightarrow (\varphi > \psi_1 \wedge \psi_2)$;
E $\forall x(\varphi > \psi) \rightarrow (\varphi > \forall x \psi)$, if x is not free in φ;
RCEA

$$\frac{\vdash \varphi \leftrightarrow \psi}{\vdash (\varphi > \chi) \leftrightarrow (\psi > \chi)}$$

RCM

$$\frac{\vdash \varphi \rightarrow \psi}{\vdash (\chi > \varphi) \rightarrow (\chi > \psi)}$$

Cond is complete (see Morreau (1992)).

The resemblance between *Bin* and *Cond* is rather remarkable, and one can ask oneself if the two representations of defaults are not identical.

In Morreau (1992) generics are indeed informally understood as quantifying over normal (typical) individuals [1]. However, representing the sentence 'Normally, φ's are ψ's' by $\forall x(\varphi(x) > \psi(x))$ involves quantifying both over individuals and worlds and it is not obvious that this can be reduced to quantifying only over individuals.

A natural idea is to establish the connection between normal worlds and typical individuals as follows: call a world φ-normal, if it contains only φ-typical individuals (analogously, call an object a typical φ if it has the property φ in all φ-normal worlds). This works only if conditionals (quantifiers) are not iterated:

Theorem 4. *If φ and ψ are first-order formulas, then*

$$Bin \vdash \Pi x(\varphi, \psi) \;\Leftrightarrow\; Cond \vdash \forall x(\varphi > \psi).$$

Proof. The theorem easily follows from the two facts:
Fact 1. If φ, ψ are first-order formulas, $Cond \vdash \forall x(\varphi > \psi) \;\Leftrightarrow\; FOL \vdash \forall x(\varphi \rightarrow \psi)$;
Fact 2. If φ, ψ are first-order formulas, $Bin \vdash \Pi x(\varphi, \psi) \;\Leftrightarrow\; FOL \vdash \forall x(\varphi \rightarrow \psi)$.
In both cases, the direction from right to left is obvious. Assume that $FOL \nvdash \forall x(\varphi \rightarrow \psi)$. Then there is a first-order model M and an object d such that $M \models \varphi[d]$ and $M \not\models \psi[d]$. Let M' be a model for *Cond* consisting of just one world w, corresponding to M, and $S(w, [\varphi[d]]) = \{w\}$. Then $M', w \not\models \forall x(\varphi(x) > \psi(x))$. Therefore $Cond \nvdash \forall x(\varphi(x) > \psi(x))$. Analogously, let M" be a model for *Bin* with the same domain and interpretation function as that of M, and $\{a : M" \models \varphi(x/a)\}_T = \{a : M" \models \varphi(x/a)\}$. Then $M" \not\models \Pi x(\varphi, \psi)$ and $Bin \nvdash \Pi x(\varphi, \psi)$. $\square$

[1] We assume for the present purposes that 'normal' and 'typical' have more or less the same meaning: a normal (typical) φ is an object which has all the properties one would expect from a φ-object without having any specific information about this very object. In particular, we are not going to make a distinction between 'normal' as 'average' and 'typical' as 'having all the features of its kind in the most condensed form'.

If φ or ψ contain conditionals / generalized quantifiers the statement does not hold: for example,

$$Bin \vdash \Pi x(\varphi(x), \Pi y(\psi(y), \varphi(x))),$$

but

$$Cond \not\vdash \forall x(\varphi(x) > \forall y(\psi(y) > \varphi(x))).$$

It is however possible to find a translation from $Cond$ into a two-sorted binary quantifier language, where conditionals of the form $\varphi(x) > \psi(x)$ are translated modulo some tricks as $\Pi w(\varphi(x, w), \psi(x, w))$. To make the embedding faithful, the extensionality axiom of Bin has to be restricted to the formulas with the same world variables. The proof can be found in Alechina (1993).

4 Applications to defeasible reasoning

The problem of defining defeasible inference (from the knowledge of defeasible generalizations to the knowledge of fact) in the language of binary quantifiers reminds of the so-called direct inference problem in probabilistic logic (from the knowledge of objective probabilities to subjective degrees of belief in individual events).

In Bacchus (1990) the latter problem was solved as follows. Let Γ be a finite knowledge base, and $\varphi(a)$ a statement concerning an object a [2]. The subjective degree of belief in $\varphi(a)$ given Γ, $prob(\varphi(a)|\Gamma)$, is computed from the objective probability $P(\{x : \varphi(x/a)\}|\{x : \Gamma(x/a)\})$. Our definition of defeasible inference is analogous.

Definition 5. Let Γ be a finite set of formulas, and $\varphi(x)$ a formula with one free variable x. Then $\varphi(a)$ defeasibly follows from Γ in Bin (in symbols $\Gamma \rhd \varphi(a)$), if $\Gamma \vdash_{Bin} \Pi x(\Gamma(x/a), \varphi(x/a))$.

The following principles of defeasible reasoning are widely accepted as the desirable ones:

Defeasible modus ponens (the names come from Morreau (1992)):
Normally, A's are B's; *a is A* $\rhd$ (It is plausible that) *a is B*.

Specificity *Normally, A's are B's*; *Normally, C's are not B's*; *All C's are A's*; *a is A*; *a is C* $\rhd$ (It is plausible that) *a is not B*.

Irrelevance *Normally, A's are B's*; *a is A*; *a is C* $\rhd$ (It is plausible that) *a is B* (if C is irrelevant or positive with respect to B).
Of course, if C is not irrelevant (for example, $C = \neg B$), the inference should not go through.

[2] We assume here that φ is a monadic property. For dealing with polyadic properties one can introduce a polyadic quantifier in an obvious way: quantifying over typical tuples of objects.

Those principles can be used as criteria of adequacy for the systems of defeasible reasoning.

One can easily see that our definition of defeasible inference for *Bin* makes the first two principles valid.

Theorem 6. *In Bin, defeasible modus ponens and specificity are valid, i.e.*

$\Pi x(B(x), F(x)), B(a) \rhd F(a)$;
$\Pi x(B(x), F(x)), \Pi x(P(x), \neg F(x)), \forall x(P(x) \to B(x)), B(a), P(a) \rhd \neg F(a)$.

Proof. The first statement is trivial. As for specificity, we need to show that

$$\Gamma \vdash \Pi x(\Gamma(x/a), \neg F(x/a)),$$

where $\Gamma(x/a)$ is

$$\Pi x(B(x), F(x)) \wedge \Pi x(P(x), \neg F(x)) \wedge \forall x(P(x) \to F(x)) \wedge B(x) \wedge P(x).$$

Clearly,

$$\Gamma \vdash \forall x(\Pi x(B(x), F(x)) \wedge \Pi x(P(x), \neg F(x)) \wedge \forall x(P(x) \to F(x)) \wedge B(x) \wedge P(x) \leftrightarrow$$
$$\leftrightarrow P(x)).$$

By extensionality and

$$\Gamma \vdash \Pi x(P(x), \neg F(x)),$$

we have

$$\Gamma \vdash \Pi x(\Gamma(x/a), \neg F(x/a)).$$

□

Note that above we essentially made use of the fact that

$$\forall x(P(x) \to B(x)) \vdash \forall x(B(x) \wedge P(x) \leftrightarrow P(x)).$$

Adding to Γ one more formula containing a, for example $Y(a)$, would block the inference, since

$$\Gamma \wedge Y(a) \nvdash \forall x(\Gamma(x/a) \wedge Y(x) \leftrightarrow P(x)).$$

Therefore, irrelevance in general is not a valid principle: we cannot justify the inference from 'Generally, birds fly', 'Tweety is a bird' *and* 'Tweety is yellow' to 'Tweety flies'.

However, there is one important special case of irrelevance which presents a difficulty for the minimal models approach but for this logic. Namely, if $\Gamma \rhd \varphi(a)$ and we add to Γ some irrelevant information Δ which does not contain a free, $\Gamma \cup \Delta \rhd \varphi(a)$, since

$$\Gamma, \Delta \vdash \forall x(\Gamma(x/a) \wedge \Delta(x/a) \leftrightarrow \Gamma(x/a)).$$

Therefore if irrelevant information is of the form 'There is a non-flying penguin' ($\exists x \neg F(x)$) or of the form 'Sam does not fly' ($\neg F(b)$), the reasoning goes thorough: we still are able to derive that Tweety flies.

5 Conclusion

We have seen that a very simple logic is suitable for representing commonsense generalizations if those are understood as talking about typical objects. Inside this logic one can perform some defeasible reasoning (defeasible modus ponens and specificity are valid), but it is in general sensitive to adding irrelevant information. Probably the only way to solve the latter problem is to use meta-theoretic arguments as it is done in other non-monotonic logics.

Comparing the properties of $\rhd$ above with the postulates of Kraus, Lehmann and Magidor (1990) shows that $\rhd$ is a very weak consequence relation. Namely, $\rhd$ satisfies Left Logical Equivalence, Right Weakening, Reflexivity, but does not satisfy Cumulativity postulates:

Cut $A \wedge B \rhd C,\ A \rhd B \Longrightarrow A \rhd C$
Cautious Monotony $A \rhd B,\ A \rhd C \Longrightarrow A \wedge B \rhd C$.

These postulates require an additional constraint on T: if all typical A's are B's ($A_T \subseteq B$), then typical A's are the same as typical $A \wedge B$'s ($A_T = (A \wedge B)_T$).

References

Alechina, N.: Binary quantifiers and relational semantics. ILLC Report LP-93-13 (1993), ILLC, Univ. of Amsterdam.

Bacchus, F.: Representing and reasoning with probabilistic knowledge. A logical approach to probabilities. Cambridge (Mass.) - London, The MIT Press (1990).

Badaloni, S., Zanardo, A.: A high plausibility logic for reasoning with incomplete time-sensitive knowledge. LADSEB-CNR (Italian National Research Council, Institute for Investigations in System Dynamics and Bioengineering), Int.Report 02/90 (1990).

Badaloni, S., Zanardo, A.: Typicality for plausible reasoning. In: Trends in Artificial Intelligence. Lecture Notes in Computer Science, Springer Verlag, (1991) 470–474.

Delgrande, J. P.: An approach to default reasoning based on a first-order conditional logic: Revised report. Artificial Intelligence **36**, (1991) 63–90.

Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models, and cumulative logics. Artificial Intelligence **44** (1990) 167 – 207.

Morreau, M. Conditionals in Philosophy and Artificial Intelligence. PhD Thesis, Univ. of Amsterdam, Amsterdam (1992).

This article was processed using the LaTeX macro package with LLNCS style