Semantics for Default Rules

Jae-Il Yeom[*†]
Hongik University

Jae-Il Yeom. 2000. Semantics for Default Rules. *Language and Information 4.2*, 69–92. It is well-known that default rules require a nonmonotonic logic. Veltman proposed one dynamic theory which interprets default rules in such a way that correct inferences can be made at each information state. But his theory has some problems. First, his theory excludes the possibility that a default rule can be true or false. Second, his representation of an information state makes it difficult to interpret a default rule embedded in another sentence. Third, the notion of a frame which is introduced in the interpretation of a default rule and the adjustment of inferential expectation has a more complex structure than is necessary. In this paper, I propose a truth-conditional theory for default rules in which the meaning of a default rule is defined as a truth-condition in a possible world and which assumes a simpler structure of a frame. This makes it possible to interpret a default rule embedded in a sentence. A dynamic theory for default rules is also proposed for correct inferences based on default rules. **(Hongik University)**

1. Introduction

Veltman (1995) proposes a dynamic theory for generic sentences which is basically concerned with how to represent an agent's knowledge of the (default) rules that he or she thinks hold in the actual world and how to adjust the agent's expectations as his or her knowledge increases.[1] An agent's expectations are based on his or her knowledge of the rules as well as the factual knowledge of the world he or she lives in. When we know that lions *normally* have four legs, and that Simba is a lion, we will expect that Simba *presumably* has four legs. It is, however, well-known that such expectations change as we get more factual information. As soon as we know that Simba has lost one leg due to an accident, we take back the inference. This shows that inferences from generic statements are defeasible. Such reasoning in the inferences is called defeasible Modus Ponens.

It has been observed that reasoning with default rules shows various other properties. Default logic follows the pattern of Modus Tollens, though it is defeasible.

(1) Defeasible Modus Tollens:
Students are (normally) unemployed.
Dick is employed.
―――――――――――――――
Dick is (presumably) not a student.

* I would like to thank two anonymous reviewers for their detailed helpful comments and suggestions, and for helping me realize how sloppy my explanation is. They are not responsible for remaining errors or inadequacies.
† 121-791 서울시 마포구 상수동 72-1 홍익대학교 영어영문과, E-mail: jiyeom@wow.hongik.ac.kr
1. Generic sentences can be grouped into two. Krifka Krifka (1987) called them i-generics and d-generics. D-generics are statements about properties which cannot be attributed to an individual, like *extinct*. On the other hand, i-generics are statements about properties which could be attributed to both individuals and kinds. In this paper, we are only concerned with i-generics.

The expectation can be canceled if the information is added that Dick is a student.

When defeasible Modus Ponens and defeasible Modus Tollens are applicable, defeasible Modus Ponens wins.

(2) Students are (normally) adults.
Adults are (normally) not students.
Dick is a student.
―――――――――――――――
Dick is (presumably) an adult.

But for the first sentence, we would expect that Dick is not an adult, by way of defeasible Modus Tollens. But from defeasible Modus Ponens with the first sentence, we expect that Dick is not a student. This defeats defeasible Modus Tollens.

A default rule allows an exception rule, and not every rule plays a role in making an inference.

(3) Exception Rule:
Birds fly.
Ostriches do not fly.
Ostriches are birds.
Tweety is an ostrich.
―――――――――――――――
Tweety (presumably) does not fly.

Irrelevant Information:
Lions are dangerous.
Simba is young.
Simba is a lion.
―――――――――――――――
Simba is dangerous.

The rule that ostriches do not fly is an exception rule to the rule that birds fly. The information that Simba is young plays no role in making the inference that Simba is presumably dangerous.

Normality has degrees, and individuals can be normal in different degrees.

(4) Graded Normality:
Students are adults.
Students are not employed.
Dick is a student.
Dick is employed.
―――――――――――――――
Dick is an adult.

Here Dick is not completely normal because he does not conform to the second rule. Without the fact that he is not employed, he could be absolutely normal.

A default rule plays a role in making inferences when it is compatible with factual information. We also have seen that there are interactions between default rules. This means that inferences change as more information obtains. In order to make a model for such defeasible inferences, we need a nonmonotonic logic. Veltman proposes one nonmonotonic logic. In this paper, I am going to discuss some problems with his theory, and propose a new analysis of default rules which can avoid the problems.

2. Veltman's Theory

Veltman is interested in the logical properties of two operators, *normally* and *presumably* within update semantics. The first operator indicates that the rest of the sentence is a (default) rule, and the second tells us that the rest of the sentence is an inference (or an expectation) based on some rules in an agent's knowledge state. In his theory, an information state σ is a pair $\langle \pi, s \rangle$. Here s is a subset of the set of possible worlds: it represents an agent's factual knowledge about the world. When a new fact is learnt, it excludes all possible worlds in s which are not compatible with the fact, and we get a new set of possible worlds $s' \subseteq s$. The first element of the pair π represents the agent's knowledge of the (default) rules. It changes as the agent learns new rules. A frame π on W is defined as follows:

Definition 1

Let W be the set of possible worlds. π is an (expectation) frame on W iff π is a function that assigns to every subset d of W a pattern $\pi(d)$ on d.

ε_d is an (expectation) pattern on d iff ε_d is a preordering of d: that is,

(i) ε_d is reflexive: for every $i \in d$, $\langle i, i \rangle \in \varepsilon_d$;

(ii) ε_d is transitive: for every i, j, $k \in d$, if $\langle i, j \rangle \in \varepsilon_d$ and $\langle j, k \rangle \in \varepsilon_d$, then $\langle i, k \rangle \in \varepsilon_d$. $\langle i, j \rangle \in \varepsilon_d$ iff i conforms to every proposition in a domain d that j conforms to'.

Suppose that d is a domain to which a newly-learnt rule is to be added. A rule changes the ordering of possible worlds in d by eliminating some pairs of possible worlds in d. For example, if a rule '$\phi \succ \psi$' (which reads, 'ϕ normally implies ψ') is learnt, the pattern on the domain $||\phi||$ changes so that possible worlds in which ψ holds take precedence over those in which ψ does not hold. This is defined as follows:

Definition 2

Let π be a frame and $d \subseteq W$.

$\langle \pi, s \rangle [\phi \succ \psi] = \langle \pi', s \rangle$, where

$\pi'(d) = \{\langle i, j \rangle \in \pi(d) | \text{if } j \in ||\psi||, \text{then } i \in ||\psi||\}$ if $d = ||\phi||$[2]

$\pi'(d) = \pi(d)$ otherwise.

A default rule is a sentence with a binary conditional connection '$\succ$.' '$\phi \succ \psi$' is understood as saying that if ϕ holds, then it is normal that ψ holds. The antecedent clause ϕ of a default rule '$\phi \succ \psi$' provides the domain $||\phi||$ in which the consequent clause ψ of the rule changes the ordering of possible worlds. In the domain $||\phi||$, the possible worlds in $||\psi||$ are taken to be more normal than those not in $||\psi||$. This is expressed by eliminating the pairs of the form $\langle j, i \rangle$, where $j \notin ||\psi||$ and $i \in ||\psi||$, from $\pi(||\phi||)$. The same rule does not change the ordering of possible worlds in other domains. Whether or not this rule applies to other domains depends on the rules which hold in the domains, and we do not know what rules will be added to them later. So a newly-learnt rule in d is not applied to other domains at this point. Applying a rule in one domain to others is necessary for making inferences from already given rules, which is a separate process. This is discussed in Definition 7.

Two incompatible rules lead to an absurd state. The notion of absurdity in a frame can be defined as consistency, which is in turn defined in terms of normal worlds:

Definition 3

Let π be a frame on W. For a domain $d \subseteq W$ and a possible world $i \in d$,

2. Here I use the truth-conditional definition of interpretation function $|| \cdot ||$ for convenience, which takes a sentence as input and yields a set of possible worlds in which the sentence is true. This is what Veltman actually does, so I will not suggest how to eliminate it.

(i) i is a (absolutely) normal world in d iff for every domain $d' \subseteq d$ such that $i \in d'$, i conforms to every rule in d';

(ii) $norm_\pi(d)$ is the set of all normal worlds in d.

(iii) π is coherent iff for every nonempty $d \subseteq W$, $norm_\pi(d) \neq \emptyset$

If a frame π is to be coherent, for every domain $d \subseteq W$ there is at least one possible world which is compatible with every rule in the domain itself d and in every smaller domain than d. This is illustrated in the following.

(5) If John is at home, normally he listens to music.
If John is at home and is busy, normally he does not listen to music.

The first sentence conveys a more general rule than the second. That is, the second is an exception rule. This exception is acceptable because there are still possible worlds in which John is at home and is not busy and in which he normally listens to music. But the following sentence excludes the normal worlds in the previous frame, and gives rise to an incoherent frame.

(6) If John is at home and is not busy, normally he does not listen to music.

The domain in which John is at home has no normal world.

In this theory, a new rule changes the pattern on the relevant domain into a more refined one in the sense that even the possible worlds which have the same degree of normality in the previous pattern are furthermore differentiated with respect to the new rule. Suppose that $p = \{1,2,3,4\}$, $q = \{3,4,5,6\}$, and $r = \{1,3,5,7\}$, where each number is a possible world. When '$p \succ q$' is learnt, the domain p is divided into two groups: A={3,4} and B={1,2}. A possible world in A is more normal than any possible world in B. When a new rule '$p \succ r$' is learnt, the two groups become four groups:

```
(7)            {3}
               / \
            {1}   {4}
               \ /
               {2}
```

A possible world 3 is the most normal. Possible worlds 1 and 4 are less normal than 3 by one rule, more normal than 2 by one rule, but cannot be compared with each other. We have seen that a new rule in a domain divides the possible worlds in that domain into more groups. In this sense the pattern is more refined. Note that the addition of a new rule just refines the pattern on the relevant domain. In order to understand what it really means, it is necessary to see how an inference changes after the rule is added.

The direct effect of a default rule in Veltman's theory is to change a pattern on the relevant domain. But we know that when we make an inference from default rules, we need to consider the interactions between them, together with the current factual information. As shown above, a rule may, or may not, affect our expectation according to whether there is a rule incompatible with that rule. And the crucial domain is the current *factual* information state. So in order to make a correct inference, we need to decide which rule applies to the domain of the current *factual* information state. This is a rather complex process. For the better understanding of the process, I will give a whole picture of the process first, and give the details in Veltman's terms.

A rule changes a pattern on a domain, as given in Definition 2. But the rule does not have to apply only to that domain. When we learn that birds fly, we naturally think that swallows fly. For this to come out, the rule that birds fly applies to other domains to which the rule is applicable. Now we need to know when a rule in a domain applies to another domain. The idea is that a rule is applicable to a domain if there are no rules incompatible with it which would apply to the same domain without the rule. When we make an inference under the current factual information state s, we consider every rule that is applicable to a domain $d \subseteq s$. If a rule e in a domain d is applicable to the domain $s \cap d$, the possible worlds in s that are not in d (= the possible worlds which are not subject to the rule) and the possible worlds that are in $s \cap e$ (= the possible worlds which conform to the rule) are more normal than the possible worlds which are subject to, but do not conform to, the rule.[3] Every possible world in s is compared with respect to every rule that applies to a domain $d \subseteq s$, and we get a new resulting pattern on s. An inference statement is tested with respect to the new pattern.

Veltman discusses the process with some supplementary notions.

Definition 4

Let e be a rule in $\pi(d)$. Then e ***applies to*** d' iff $d' \subseteq d$ and for every $d'' \supseteq d'$ such that $norm_\pi(d'') \subseteq d - e$, it holds that $d'' \subseteq d - e$.

i is subject to e iff $i \in s \cap d$ and e applies to $s \cap d$.

i is an accidental exception to a rule e iff i is subject to e and $i \notin e$.

$\langle i, j\rangle \in \varepsilon_\sigma$ iff i, $j \in s$, and j is an accidental exception to every rule to which i is an accidental exception.

Here ε_σ is a resulting pattern on s we get after applying every rule in π to lower domains to which they are applicable, and thus to s. A rule e in a domain d changes the ordering of possible worlds in s if it can apply to every domain $d'' \supseteq d \cap s$ without making $norm_\pi(d'')$ the empty set. If there is a domain $d'' \supseteq d \cap s$ in which the set of normal worlds becomes the empty set, it implies that there is an exceptional rule to e in some lower domain than d. Note that an exceptional rule takes over a more general rule. But if there is no possible world in a domain $d'' \supseteq d'$ that conforms to a rule, then it means that the rule has no discriminating effect on the possible worlds in the domain. In this case, we can assume that the rule vacuously applies to the domain, so it does not block the rule e from applying to d'. Now when we decide ordering between possible worlds, the notion of accidental exception divides possible worlds in s into two groups; the possible worlds in which are subject to e but do not conform to it and those which are not subject to it or which are subject to and conform to the rule. The former are $s \cap (d - e)$ and the latter $(s - d) \cup (s \cap e)$. A possible world w in the latter is taken to be more normal than one in the former. This implies that '$\phi \succ \psi$' has the effect of taking as more normal those possible worlds in which '$\neg\phi \vee \psi$' is true.

Considering every rule that affects the ordering of possible worlds in s, we can test an inference as to whether it is a correct one in the current information state. The adverb *presumably* is often used in inference statements. It is one-place operator which takes a sentence as its argument. The operator is translated into $\triangleright$ below.

Definition 5

Let $\sigma = \langle \pi, s\rangle$. $\sigma[\triangleright\phi] = \begin{cases} \sigma \text{ if } opt_\sigma||\phi|| = opt_\sigma \\ 0, \text{otherwise} \end{cases}$

$i \in opt_\sigma$ iff $i \in s$ and $\langle i, j\rangle \in \varepsilon_\sigma$ for every $j \in s$ such that $\langle j, i\rangle \in \varepsilon_\sigma$.

3. Here 'a rule e in a domain d' corresponds to a rule $\phi \succ \psi$, where $||\phi|| = d$ and $||\psi|| = e$.

opt_σ is the set of all optimal worlds in $\sigma = \langle\pi, s\rangle$. A possible world is an optimal world if it is in s and it is the most normal with respect to every rule that applies within s. A sentence *'Presumably ϕ'* roughly means that ϕ holds in every optimal world in the current information state.

I briefly introduced Veltman's analysis of default rules. To summarize, he was basically concerned with how to explain how inferences are made on the base of learnt rules at each information state. Basically we need two steps to analyze default rules. One is the process in which a rule is added to the relevant domain. This is rather a direct addition of a rule. In this process, if there is inconsistency in a frame, the information state becomes an absurd state. The other process is the one in which rules in a domain affect other domains. In this step, if a rule causes inconsistency in a domain, the rule does not apply to the domain. All inference mechanism comes from the second process.

3. Negation of a Default Rule in Veltman's Theory

One problem with Veltman's theory is that a statement '$\phi \succ \psi$' changes the pattern on the domain $||\phi||$. A negative rule '$\phi \succ \neg\psi$' does not cause any problem: among the possible worlds in $||\phi||$, it makes possible worlds in which ψ holds less normal than those in $||\neg\psi||$. A problem arises when the negation of a rule is interpreted. What does it mean in Veltman's theory? It must mean that the affirmative counterpart must not hold. But Veltman's theory cannot express such a meaning. The problem comes from the fact that each domain has only one pattern and that the role of a default rule is simply to change the pattern. In order to interpret the negation of a rule correctly, a domain needs to have a more complex structure. In this section, I will show that the problem can be solved by assuming that a domain has a set of patterns and that a default rule in a domain eliminates some patterns on the domain which are not compatible with the rule.

An information state consists of two components, π and s, as in Veltman's theory. The first conveys the knowledge of rules learnt by an agent, and the second the factual knowledge of the actual world.

Definition 6

Let W be the set of possible worlds. Then σ is an information state iff $\sigma = \langle\pi, s\rangle$, where π is a frame on W and s is a subset of W.

The minimal information state is $\langle\pi, W\rangle$, where π is a frame such that for every $d \subseteq W$, $\pi(d)$ is the set of all possible patterns on d.

In the minimal information state, there is no factual information about the actual world. This is represented as the whole set of possible worlds W. And there is no knowledge of any default rule. It leaves every expectation pattern in the frame without any ordering between possible worlds. This is represented by the set of all possible patterns. In general, a frame has the form of the following:

Definition 7

Let π be an (expectation) frame on W. π is a function from $\wp(W)$ to $\{\{\varepsilon \subseteq d \times d \mid \varepsilon \text{ is reflexive and transitive}\} \mid d \in \wp(W)\}$.

That is, for every domain $d \subseteq W$, there is a set of patterns$\pi(d)$ which is a subset of $\{\varepsilon \subseteq d \times d \mid \varepsilon \text{ is reflexive and transitive}\}$. Each domain has a different set of patterns, and a set of patterns indicates which possible world is at least as normal as which possible world.

Definition 8

A possible world i is at least as normal as a possible world j in the domain d with respect to the rules in d ($i \geq_d j$) iff for every pattern ε in $\pi(d)$, if $\langle j,i\rangle \in \varepsilon$, then $\langle i,j\rangle \in \varepsilon$.

If a pattern in a domain contains $\langle i,j\rangle$, but not $\langle j,i\rangle$, and if another pattern in the same domain contains $\langle j,i\rangle$, but not $\langle i,j\rangle$, it is not determined which of the two possible worlds is more normal. Each domain has a set of patterns, and the ordering of possible worlds in one domain is different from that in another. If different domains with the same two possible worlds contain two different default rules, and if one of the rules makes one possible world more normal and the other rule makes the other possible world more normal, then both possible worlds are not really normal worlds. For this reason, a normal world in a domain is defined with respect to the rules in every smaller domain as well as the rules in that domain.

Definition 9

Let π be a frame on W. For a domain $d \subseteq W$ and a possible world $i \in d$,

(i) i is a normal world in d iff for every $d' \subseteq d$ such that $i \in d'$, $i \geq_d j$ for every $j \in d'$;

(ii) $norm_\pi(d)$ is the set of all normal worlds in d.

The notion of normal world is defined in terms of the comparison of normality, which is defined with respect to every pattern in a domain, as given in Definition 3.

When a default rule is learnt, it will change the frame. This is defined by the interpretation function. The interpretation function $[[\cdot]]$ is defined as follows.

Definition 10

Let $\sigma = \langle \pi, s\rangle$.

$\sigma[[p]] = \langle \pi, \{w \in s \mid w \in F(p)\}\rangle$, where F is a function from non-logical constants to their denotations.[4]

$\sigma[[\neg\phi]] = \sigma \sim [[\phi]]$

$\langle \pi, s\rangle \sim \langle \pi', s'\rangle = \langle \pi'', s''\rangle$, where

$$\begin{cases} \text{for every } d \subseteq W \text{ such that } \pi(d) = \pi'(d),\ \pi''(d) = \pi(d) \\ \text{for every } d \subseteq W \text{ such that } \pi(d) \neq \pi'(d),\ \pi''(d) = \pi(d) \setminus \pi'(d) \\ \text{if } s = s', \text{ then } s'' = s \\ \text{if } s \neq s', \text{ then } s'' = s \setminus s' \end{cases}$$

$\sigma[[\phi \wedge \psi]] = \sigma[[\phi]][[\psi]]$

$\sigma[[\phi \succ \psi]] = \langle \pi', s\rangle$, where

if $d \neq \{w \in W \mid \phi \text{ holds in } w\}$, then $\pi'(d) = \pi(d)$.[5]

if $d = \{w \in W \mid \phi \text{ holds in } w\}$, then $\pi'(d) = \{\varepsilon \in \pi \mid \varepsilon[\psi] = \varepsilon\}$.

4. In Veltman's theory, a possible world is a set of propositions which hold in the possible world. For this reason he says that $F(p) \in w$, not $w \in F(w)$. One reason I can think of is that he assumes that a rule in a domain can be retrieved even after it is interpreted and added to a frame. Update is eliminative, and normally a sentence which is interpreted eliminatively cannot be retrieved. But if a possible world consists of propositions which hold in that possible world, we can retrive a rule which has already been interpreted.

5. The use of the set $\{w \in W \mid \phi \text{ holds in } w\}$ is only for convenience. For the consistent use of the interpretation function, it should be something like $\sigma[[\phi]]_2$. ($\langle\pi, s\rangle_1 = \pi$ $\langle\pi, s\rangle_2 = s$).

In the interpretation of a default rule, a new interpretation function $[\cdot]$ is introduced.[6] It changes the ordering of possible worlds in the domain to which the default rule is added. The interpretation function $[\cdot]$ is defined as follows:

Definition 11

Let ε be an (expectation) pattern on a domain d.
$\varepsilon[p] = \{\langle i,j\rangle \in \varepsilon \mid j \notin F(p) \vee i \in F(p)\}$
$\varepsilon[\neg\phi] = \{\langle i,j\rangle \in \varepsilon \mid j \in \{k \mid \exists l[\langle l,k\rangle \in (\varepsilon \setminus \varepsilon[\phi])]\} \vee i \in \{l \mid \exists k[\langle l,k\rangle \in (\varepsilon \setminus \varepsilon[\phi])]\}\}$
$\varepsilon[\phi \wedge \psi] = \varepsilon[\phi][\psi]$

If a pattern in a domain of a frame is updated with an atomic sentence, all possible worlds in which the sentence is true at least as normal as those in which it is not true. This eliminates all pairs of possible worlds $\langle i,j\rangle$ such that the sentence is true in j and false in i. When ε is updated with a negative sentence $\neg\phi$, it must be defined based on $\varepsilon[\phi]$ so that the interpretation could be done compositionally. By subtracting $\varepsilon[\phi]$ from ε, we get the pairs of possible worlds $\langle i,j\rangle$ where ϕ is true in j and false in i. Thus $\{j \mid \exists i[\langle i,j\rangle \in \varepsilon\setminus\varepsilon[\phi]\}$ is the set of possible worlds in which ϕ is true, and $\{i \mid \exists j[\langle i,j\rangle \in \varepsilon \setminus \varepsilon[\phi]\}$ is the set of possible worlds in which ϕ is false. Then the negative sentence $\neg\phi$ is false in the former set, and true in the latter. It is clear that the two sets are exclusive to each other: no possible world belongs to both sets. It is necessary to show that the union of the two sets is W.

If a default rule '$\phi \succ \psi$' is any different from every rule in the current information state, it will tell us for every possible world w whether ψ is true in w. Suppose that there is a possible world i in which a new rule ψ is false and in which every rule in the current information state is true. Then if $\langle i,j\rangle$ remains in the information state after the update of the information state with the new rule, the new rule is false in j. From this, we can get the set of possible worlds in which the rule is false. $\{i \mid \exists j[\langle i,j\rangle \in \varepsilon \setminus \varepsilon[\phi]\}$ is what we get. If the pair $\langle i,j\rangle$ is excluded from the information state, the new rule is true in j. From this, we can get the set of possible worlds in which the new rule is true. $\{j \mid \exists i[\langle i,j\rangle \in \varepsilon \setminus \varepsilon[\phi]\}$ is that set. The two sets are complementary. So the union of the two sets is W. If there is no possible world in which the new rule is false and every rule in the current information state is true, the new rule conveys no new information. When the current information state ε contains the information conveyed by ϕ, $\varepsilon \setminus \varepsilon[\phi]$ is the empty set. In this case, both $\{i \mid \exists j[\langle i,j\rangle \in \varepsilon \setminus \varepsilon[\phi]\}$ and $\{j \mid \exists i[\langle i,j\rangle \in \varepsilon \setminus \varepsilon[\phi]\}$ are the empty set. Then $\varepsilon[\neg\phi]$ results in the empty set. It means that the same domain has two incompatible rules, which makes the frame incoherent. In other words, the resulting information state is the absurd state.

Now we can show how a rule updates the information state. Suppose that W is the set $\{w_0, w_1, w_2, w_3\}$. The proposition p is true in w_1 and w_3, and the proposition q is true in w_2 and w_3. Now suppose that the current information state is as follows:

(8) the current information state $\sigma = \langle\pi, s\rangle$, where $\pi(\{w_1, w_3\}) = \{$
$\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\},$
$\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle\},$
$\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_3, w_1\rangle\},$
$\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle, \langle w_3, w_1\rangle\}\}$

6. The reason why we need two interpretation functions is that the inputs for the two functions are different. The input of the interpretation function $[[\cdot]]$ has the form of $\langle\pi, s\rangle$ while the input for $[\cdot]$ is the set of pairs of possible worlds.

When a sentence '$p \succ q$' is uttered, it only changes the domain $\{w_1, w_3\}$. The update rule says that $\pi(\{w_1, w_3\})$ changes as follows:

(9) $\pi'(\{w_1, w_3\}) = \{\varepsilon \in \pi(\{w_1, w_3\}) \mid \varepsilon[q] = \varepsilon\}$

$\pi(\{w_1, w_3\})$ has four patterns. Each pattern is updated with the sentence q by the interpretation function $[\cdot]$. If a pattern does not change by the update, it remains in the frame. Otherwise, it is excluded. The expectation patterns $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\}$, and $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_3, w_1\rangle\}$ do not change by the rule q, but the expectation patterns $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle\}$ and $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle, \langle w_3, w_1\rangle\}$ change: $\langle w_1, w_3\rangle$ is excluded from the pattern because w_3 is in $F(q)$ while w_1 is not. So these patterns are excluded from the frame, and the resulting information state becomes the following:

(10) the current information state $\sigma = \langle \pi', s\rangle$, where $\pi'(\{w_1, w_3\}) = \{ \{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\}, \{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_3, w_1\rangle\}\}$

This shows that a default rule is stable. Once a rule selects only the patterns which are compatible with the rule, the new frame allows no rule that is not compatible with the rule.

Now suppose that a sentence '$\neg(p \succ q)$' is uttered assuming the same current information state as (8).

(11) $\langle \pi, s\rangle || \neg(p \succ q) || = \langle \pi, s\rangle \sim \langle \pi, s\rangle || p \succ q ||$
$= \langle \pi', s\rangle$, where $\pi'(\{w_1, w_3\}) = \{\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle\}, \{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle, \langle w_3, w_1\rangle\}\}$

This shows that the negation operator before a default rule actually negates the relation between the antecedent and consequent clauses of the default rule.[7] The update illustrated here shows that the negation of a default rule is also stable. In the update rule, a default rule does not change an expectation pattern on a domain. It only checks whether a pattern is compatible with the rule, and incompatible patterns are excluded. This implies that once all patterns compatible with a rule are eliminated from a domain, there is no chance that a new pattern which is compatible with the rule is introduced into that domain. So the negation of a rule always holds in that domain.

The negation of a default rule is different from a negative rule. This can be shown by showing how a negative rule updates an information state. Suppose that a sentence '$p \succ \neg q$' is uttered. It updates the information state in (8) into the following.

(12) $\langle \pi, s\rangle || p \succ \neg q) || = \langle \pi', s\rangle$, where $\pi'(\{w_1, w_3\}) = \{ \{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\}, \{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle\}\}$

Only the pattern which is compatible with $\neg q$ remains in the domain of $\{w_1, w_3\}$. This is different from the result we get from the sentence '$\neg(p \succ q)$'. The resulting set of patterns is the subset of the set of patterns we get from $\neg(p \succ q)$. This implies that the negative rule is stronger than the negation of the corresponding affirmative rule. This is intuitively correct. Consider the following two sentences.

(13) Normally it does not rain in April.
It is not the case that it normally rains in April.

7. This is a quite natural result. We have assumed that a rule has the form of '$\phi \succ \psi$' for convenience. But the operator $\succ$ is relational, so it can be written as '$\succ (\phi, \psi)$'. So the negation of a rule $\neg \succ (\phi, \psi)$ negates whatever relation there is between the two prpositions ϕ and ψ.

The first sentence is stronger than the second. So the second sentence can be followed by the negation of the first sentence.

(14) It is not the case that it normally rains in April. Neither is it the case that normally it does not rain in April.

(15) $\langle \pi, s\rangle || \neg(p \succ q) \wedge \neg(p \succ \neg q)||$
$= \langle \pi'', s\rangle$, where $\pi''(\{w_1, w_3\}) =$
$\{\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle, \langle w_3, w_1\rangle\}\}$

From the first sentence, we get (11). For the second sentence, we subtract (12) from (11).

Notice that the pattern $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\}$ conveys no information. Whether or not it is contained in a domain of a frame does not make the information state any different. Suppose that '$\neg(p \succ q)$' is followed by '$p \succ \neg q$'. Then the first sentence is weaker than the second. And so we expect to get the same result only from the second. This is not the case, however.

(16) $\langle \pi, s\rangle || \neg(p \succ q)|| || p \succ \neg q|| = \langle \pi''', s\rangle$, where $\pi'''(\{w_1, w_3\}) = \{$
$\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle, \langle w_1, w_3\rangle\}\}$

This can be compared with the result in (12). The only difference lies in the existence of $\{\langle w_1, w_1\rangle, \langle w_3, w_3\rangle\}$. The main role of a pattern in a frame is to indicate which possible world is at least as normal as which possible world. The pattern we are concerned with now does not indicate any ordering between any two possible worlds. This does no harm to the theory, and so I will leave it as it is.[8]

4. Other Problems with Veltman's Theory

In the previous section, I pointed out a problem with Veltman's interpretation of the negation of a default rule. To solve the problem, we need a more complex frame than Veltman assumes. But the problem is not the only one that requires the theory to be more complex. Consider the following sentences.

(17) Mary knows that penguins normally do not fly.

(18) People who work late nights do not wake up early.

In the first sentence, the interpretation of the verb ***know*** needs Mary's knowledge state. It is generally defined by a set of epistemic alternatives (for Mary) to a possible world in which the whole sentence is true. But I suppose that Veltman's theory needs something like a pair of a frame and a set of epistemic alternatives to a possible world for the interpretation of a default rule in Mary's belief. I do not know if the meaning of such a sentence can be defined adequately. In order to interpret the second sentence with respect to a possible world, we need the denotation of the expression *people who work late nights*. But if the meaning of part of the expression (*normally*) *work late nights* is defined only with respect to a frame, how could the denotation of the expression in question be expressed? Even if the sentence could be interpreted somehow, it would require much more difficulty than the problem with the negation of a default rule.

8. This occurs in interpreting the negation of a rule, and is a purely technical matter. To avoid this, we have a variety of ways, one of which is to use a higher-order of patternsn. But this will make the interpretation function much more complex. Another way is to add $\{\langle i, i\rangle \mid i \in d\}$ to the result of what we get from the original interpretation of the negation of a rule.

The root of these problems is that in his theory the meaning of a default rule is not defined with respect to a possible world. A rule has only a dynamic meaning of refining a pattern. Intuitively a default rule seems to have a truth value, and if a rule is true in a possible world, it is contingently true. This implies that if we need a frame to interpret a rule, the frame for a possible world should be different from the one for another possible world. If the meaning of a default rule is given with respect to a possible world, we can avoid the complexity that is required in Veltman's theory.

Veltman's theory is not quite general in another respect. A generic statement is interpreted dynamically by changing the expectation frame. But in his theory, as presented in Veltman (1995), a negative default rule is not interpreted from the dynamic interpretation of its affirmative counterpart. It is generally assumed that a negative sentence is not directly verified in a possible world. So the dynamic meaning of a negative sentence must be defined with the dynamic meaning of the affirmative counterpart. Instead, he uses the static interpretation of the negative sentence. If $\phi = \neg p$, then he first derives the set of possible worlds in which p does not hold by subtracting the set of possible worlds in which p holds from W. In this sense, his theory is not completely dynamic. And the interpretation is not completely compositional, either. In the previous section, I interpreted a negative rule dynamically through the dynamic interpretation of the affirmative counterpart.

5. A Theory Based On Truth Conditions

5.1 Truth-Conditional Theory of Default Rules

A default rule can be true or false in the actual world. This is intuitively correct. Several attempts have been made following this line of analysis, but they have not been successful in accounting for the properties of inferences based on default rules. See Asher and Morreau (1995), Carlson (1988), Carlson (1982), etc. The main purpose of this paper is to analyze default rules basically following the tradition of Lewis's (1973) and Stalnaker's (1987) analyses of conditionals and show that truth-conditional semantics can do as well as Veltman's theory even in explaining the inferential characteristics. So the theory I will pursue from now on is focused on how inferences are made on the basis of default rules within the truth-conditional semantics.

A sentence '$\phi \succ \psi$' means that where ϕ holds, normally ψ holds too. In other words, ψ is one of the things that are normally the case when ϕ holds. Based on this idea, I will define the truth-condition of a default rule. For the interpretation of default rules, we need to define a frame which, for a proposition, tells us exactly what is normally the case together with that proposition. A frame is defined as follows:

Definition 12

A base model M for a language L is a tuple $\langle W, f, V\rangle$, where
$\langle W, f\rangle$ is a frame, and V is a function from non-logical constants of L to their denotations.

Before we define the truth-condition of a default rule '$\phi \succ \psi$', it is necessary to characterize a frame. A function f assigns to each possible world w a frame $f(w)$, which is a function from a set of sets of possible worlds to a set of sets of possible worlds.

Definition 13

A selection function f is a function from W to $[\wp(W) \rightarrow \wp(W)]$.

In relation to the interpretation of a default rule, this function can be understood like this. If a proposition p is given to the frame $f(w)$, then it yields a set of possible worlds $f(w)(p)$. The set of possible worlds can be understood as the conjunction of propositions that are normally the case when p holds.

With the idea of frame given above, the truth-condition of a default rule is quite clear. The interpretation function $||\cdot||$ is defined as follows:

Definition 14

For any base model $M = \langle W, f, V\rangle$ and any possible world w:
$||p||_{M,w} = 1$ iff $w \in V(p)$, where p is an atomic sentence.
$||\phi \wedge \psi||_{M,w} = 1$ iff $||\phi||_{M,w} = 1$ and $||\psi||_{M,w} = 1$
$||\phi \succ \psi||_{M,w} = 1$ iff $f(w)(||\phi||_{M,w}) \subseteq ||\psi||_{M,w}$.

An atomic sentence is interpreted by the function V, which gives non-logical constants their denotations. The interpretation of a conjoined sentence is determined by the interpretation of its parts. In the interpretation of a default rule, $||\phi||$ corresponds to a domain in Veltman's theory. In this theory, however, a frame does not change as we get more information, as in Veltman (1995). A frame $f(w)$ is determined by the rules which hold in w. So we do not need the dynamic notion of meaning insofar as default rules are concerned. If a default rule '$\phi \succ \psi$' is to be true, ψ must hold in the domain of $||\phi||$. This follows from the definition of a frame. $f(w)(||\phi||)$ yields a set of possible worlds in which everything normal holds when ϕ holds. If ψ is one of the normal things when ϕ holds, then ψ holds in every possible world in $f(w)(||\phi||)$.

This definition of selection function is too general to be used in the explanation of default logic. In order to explain some relationships between default rules, we need to constrain the function so that it can capture the semantic properties of default rules. The function is required to have the following properties. First, when a proposition d is given to a function $f(w)$, there is at least one possible world which is normally the case together with the proposition. Let w be a possible world with respect to which a sentence is evaluated. Then

(19) Condition 1: For every domain $d \subseteq W$, $f(w)(d) \neq \emptyset$.

This condition is quite natural: if there is a domain $d \in W$ such that $f(w)(d) = \emptyset$, it means that w is an impossible world. It is like the following case:

(20) Students are polite. Students are impolite.

The sentence 'Students are polite' is understood as 'for every individual a, if a is a student, it is normally the case that a is polite. If the sentence is true in a possible world w, for every individual a, '(a is a student) $\succ$ (a is polite)' is true in w. That is, the sentence can be taken to be the conjunction of as many sentences as students. In the discussion that follows, I consider only one simple default rule for one individual, here '(a is a student) $\succ$ (a is polite)' for an individual a. The two default rules cannot be both true in a possible world. The two incompatible rules apply to the same domain in which a is a student, and makes $f(w)(||a$ is a student$||)$ the empty set. This is excluded from the definition of possible world. But this is a weaker condition for a coherent frame. Consider the following:

(21) $p \succ q$. $(p \wedge r) \succ \neg q$. $(p \wedge \neg r) \succ \neg q$.

In no possible world can these three hold together. This shows that when we want to see if a domain has the empty set of normal worlds, we need to consider consider its smaller domains.

Definition 15

A possible world w' is a ***(absolutely) normal world*** in d in the frame from a possible world w iff $w' \in f(w)(d')$ for every domain $d' \subseteq d$ such that $w' \in d'$.

$N_{f(w)}(d)$ is the set of normal worlds in a domain d in the frame from the possible world w.

A normal world must be defined with respect to a domain because it is possible that one possible world is a normal one while it is an abnormal one in another domain. This is illustrated below.

(22) Adults are employed.
College students are not employed.

Assume that college students are all adults. If there are some college students who are unemployed, they are normal college students, but they can be taken to be abnormal adults. An abnormal adult can be a normal college student.

We can define a coherent frame as follows. If an incoherent frame were given by a world, the world would be an impossible world.

Definition 16

The frame for a possible world w, $f(w)$, is coherent iff for every domain $d \subseteq W$, $N_{f(w)}(d) \neq \emptyset$.

Every frame from a *possible* world must be coherent.

A second condition comes from the definition of the function f.

(23) Condition 2: $f(w)(p) \subseteq p$

This condition is quite clear from the definition of frame: $f(w)(p)$ is a set of possible worlds which contains everything that is normally the case when p holds. Therefore the possible worlds must be among those in which p holds.

A third condition on a frame is given between a larger domain and its two smaller domains. This condition roughly says that a rule in a smaller domain also holds in the larger domains unless the larger domains include an incompatible rule with it.

(24) Condition 3:
For every two domains d, $d' \subseteq W$, (i) $f(w)(d \cup d') = f(w)(d) \cup f(w)(d')$ or (ii) $f(w)(d \cup d') \cap f(w)(d) = \emptyset$ or (iii) $f(w)(d \cup d') \cap f(w)(d') = \emptyset$.

The first case in the condition (i) holds in a case like the following:

(25) Tomatoes normally contain vitamin C.
Apples normally contain vitamin C.
―――――――――――――――――――
Tomatoes and apples normally contain vitamin C.

If the first two rules hold in a possible world, the third one must hold in the world. Note that the rules in the three domains are compatible with each other. Consider the following case:

(26) Birds fly.
Penguins do not fly.

In this case, birds which are not penguins fly and penguins do not fly. For an individual a and a possible world w, p='a is a bird', and q= 'a is a penguin.' Then $f(w)(||p||)$ is not (a subset of) $f(w)(||p \wedge \neg q||) \cup f(w)(||p \wedge q||)$. Instead, $f(w)(||p||) \cap f(w)(||p \wedge q||) = \emptyset$ because the rule in the domain $||p||$ is incompatible with the one in $||q||$. A case like this occurs when one rule is exceptional to another rule.

Among the three options (i), (ii), and (iii) in the condition, no two options can hold at the same time. If $f(w)(d \cup d') = f(w)(d) \cup f(w)(d')$ and $f(w)(d \cup d') \cap f(w)(d) = \emptyset$, then $f(w)(d)$ must be the empty set, which is impossible due to the second condition. We can say the same thing about the case where $f(w)(d \cup d') = f(w)(d) \cup f(w)(d')$ and $f(w)(d \cup d') \cap f(w)(d') = \emptyset$. If $f(w)(d \cup d') \cap f(w)(d) = \emptyset$ and $f(w)(d \cup d') \cap f(w)(d') = \emptyset$, $N_{f(w)}(d \cup d')$ is the empty set. This makes the frame $f(w)$ incoherent. So each of the three options in the condition holds exclusively of the others.

One thing I want to mention is that in (26) we cannot conclude that birds normally are not penguins. But this example is not good for discussion because the relation between birds and penguins is taxonomic and exceptional penguins which fly are hard to imagine. Consider the example in (22) again. Assume for an individual a that p = 'a is an adult', q = 'a is a college student', and r = 'a is employed.' If the two sentences are true, $f(w)(||p||) \subseteq ||r||$ and $f(w)(||q||) \subseteq ||\neg r||$. And $f(w)(||p||) \neq f(w)(||p \wedge q||) \cup f(w)(||p \wedge \neg q||)$. In this case there may be some exceptional college students who are employed. The exceptional students can be taken to be normal adults. This means that there are students among normal adults: that is, $f(w)(||p||) \cap ||q||$ is not the empty set. So we cannot say that normal adults are not college students.

5.2 Dynamic Theory of Default Rules

In a dynamic theory of meaning, it is assumed that the meaning of a sentence is a function that updates an information state. As pointed out in the previous section, however, it is also intuitively correct that a default rule can be true or false in a possible world. In this section, I will propose a theory in which a default rule updates an information state but the update is still based on a truth-conditional notion of meaning.

An information state and a dynamic interpretation function $[\cdot]$ is defined as follows:

Definition 17

The information model based on the base model $\langle W, f, V \rangle$ is the tuple $\langle \wp(W), f, V, [\cdot] \rangle$, where $\wp(W)$ is the set of information states allowed in the model, f a selection function, and V a function from non-logical constants to their denotations.

Let L be the set of sentences in a language. The dynamic interpretation function $[\cdot] : \wp(W) \times L \rightarrow \wp(W)$ is defined as follows:

$s[p] = s \cap V(p)$, where p is an atomic sentence.

$s[\neg \phi] = s \backslash s[\phi]$

$s[\phi \wedge \psi] = s[\phi][\psi]$

$s[\phi \succ \psi] = \{w \in s \mid f(w)(W[\phi])[\psi] = f(w)(W[\phi])\}$[9]

The other interpretation rules than the one for default rules are just like those in other dynamic theories. The interpretation rule for default rules is a slight modification of the truth condition of default rules. But I cannot use the interpretation function $|| \cdot ||$ of truth-conditional semantics. Using the dynamic interpretation function $[\cdot]$, the same set

9. For every information state $s \in \wp(W)$ and every sentence $\phi \in L$: $s[\phi] = \{w \in s \mid w \in [[\phi]]\}$.
This assumes that dynamic interpretation can be defined in terms of the truth-conditional semantics. This, however, is not supported in interpreting other operators we do not consider here.

as $||\phi||$ is obtained by $W[\phi]$. The use of W implies that the semantics for default rules is essentially static. That is, the interpretation of a default rule does not change regardless of what rules are already known. Even though a frame is used in interpreting a default rule, it is given by a possible world. So a default rule changes an information state s into a subset s' which only contains possible worlds in which the rule is true.

So far we have not seen how an inference is made from default rules we have learnt. In making an inference, we use a new operator ***presumably***, which is written as $\rhd$ below. Inference is a matter of information, not a matter of truth. In order to make a prediction in the current information state s, we need to derive a set of rules which applies to s among the rules known in the current information state. A rule which holds in a domain may, or may not, apply to another domain according to whether there is a rule which is incompatible with the rule. This must be captured by the relationships between the rules in different domains. For this purpose, we need a new condition which is imposed between a domain d and another domain. This condition is not given in the previous section, because it does not specify what default rules must hold if some rules hold in a possible world. It plays a more significant role in determining which inference can be made from some default rules given. This is why I discuss this condition in this section.

In giving a new condition, it is not sufficient only to consider the relationships between the sets of normal worlds. We need to consider whether each rule in a domain applies to a certain domain. Consider the following example.

(27) Adults are employed.
Adults are responsible.
College students are not employed.

It is the first rule, not the second, that is incompatible with the third rule. In this case, the first rule does not apply to the domain in which an individual a is a college student, but the second does. Using p, q, and r as in (22), when $f(w)(||p||) \cap f(w)(||q||) = \emptyset$, no condition makes sure that the second rule applies to the domain $||q||$. For this reason, we need to specify some condition with respect to each rule, not with respect to the set of possible worlds in which a set of rules holds in a domain.

Consider another example.

(28) Graded Normality:
Birds normally fly
Birds normally have two legs.
Tweety is a bird.
Tweety does not fly.

Tweety presumably has two legs.

The domain in which Tweety is a bird contains the first two rules. But to the domain in which Tweety is a bird and does not fly, only the second rule applies. This shows that conditions between domains must be specified with respect to rules. Otherwise, it would be specified that the first two rules together should, or should not, apply to the domain in which Tweety is a bird and does not fly.

Condition 3 is not adequate in this respect, either. Consider the following example.

(29) Adults are employed.
College students are unemployed.
College students have parties on Fridays.

In this example, the second rule does not apply to the domain for the first sentence, but the third does, though both the second and third rules are the rules in the same domain. So the relationships between domains must be specified with respect to rules, not with respect to a set of possible worlds which conform to a set of rules.

For this reason, I first define the notion of a rule being in a domain.

Definition 18

A rule e is in the domain d iff[10]

(i) $f(w)(d) \cap e = f(w)(d)$,

(ii) for every domain $d' \subseteq d$, $f(w)(d') \cap e = f(w)(d')$ or $f(w)(d') \cap e = \emptyset$, and

(iii) if there is a domain $d' \supseteq d$ such that $f(w)(d') \cap e \neq f(w)(d')$, then for every domain d'' such that $d \subseteq d'' \subseteq d'$, $f(w)(d'') \cap e \neq f(w)(d'')$.

The first condition in the definition does not need explanation. The second condition says that if a rule is in a domain, it must hold in every lower domain unless an incompatible rule holds in that domain. The third condition is included just to reflect our intuition about the idea of rule. When a more general rule in a domain applies to a smaller domain, it is not natural to say that the rule is in the smaller domain. For example, if birds fly, sparrows, swallows, parrots, etc. fly. In this case, we do not say that there are four or more rules. We think that we have one general rule. If a rule is in a domain, there is no rule which is more general than the rule in a larger domain. One might say that the third condition should be like 'there is no larger domain in which $f(w)(d') \cap e = f(w)(d')$, but this is not sufficient for the following case:

(30) Adults are employed.
College students are not employed.
Poor college students are employed.

The third rule holds in a domain in which an individual is a poor college student, but still there is a domain in which the same rule holds in a larger domain in which the first rule holds. For this case, a more complex condition is necessary.

A new condition must be given on a frame in terms of rules.

(31) Condition 3: (Revised)
If a rule e is in a domain $d \subseteq W$, then for every other domain d' such that $f(w)(d') \backslash (d \backslash e) \neq f(w)(d')$, there is a domain $d'' \supseteq (d \cap d')$ such that $N_{f(w)}(d'') \cap d \subseteq d \backslash e$ and $d'' \cap d \not\subseteq d \backslash e$.

This condition says that a rule in a domain d applies to every other domain d' when the rule can apply to the domain $d \cap d'$. And a rule in a domain d can apply to a domain d' if there is no rule in a domain $d'' \supseteq (d \cap d')$ which is incompatible with the rule. As pointed out above, the frame from a possible world is always coherent. So if a rule in a domain d does not apply to a domain d', it means that there must be some domain $d'' \supseteq (d \cap d')$ which contains some rule which is incompatible with the rule in d. The reason that we consider the domains at least as large as $(d \cap d')$ is that if a rule e in d applies to a domain d', it must be able to apply to the domain $(d \cap d')$. When the rule in d applies to a domain

10. Here the term 'rule' may be confusing because it is different from its use in other places. If a rule e is in the domain d, it comes from the sentence '$\phi \succ \psi$', where $||\phi|| = d$ and $||\psi|| = e$.

d', every possible world in $f(w)(d')$ must be in e or in the complement of d. If a possible world is in $d\backslash e$, then it means that the world is abnormal with respect to the rule: the world is less normal than any possible world in the complement of $d\backslash e$. Therefore if a rule e in d applies to a domain d', $f(w)(d')\backslash(d\backslash e) = f(w)(d')$. If the rule does not apply to d', it means that there is some incompatible rule in some domain $d'' \supseteq (d \cap d')$. When the set of normal worlds in d'' is a subset of $d\backslash e$, the application of the rule in d to a domain smaller than d'' would make the set of normal worlds in d'' the empty set. This would make the frame incoherent. Therefore the rule does not apply to any domain larger than $(d \cap d')$. There is one caveat about this. When $N_{f(w)}(d'') \subseteq d\backslash e$ *and* $d'' \subseteq d\backslash e$, there is no incompatible rule with the one we want to apply. The set of normal worlds would become the empty set only because the domain itself is not compatible with the rule. In this case, the rule does not apply to that domain, but it can apply to larger domains than that domain.

This is illustrated in (1–2), which is repeated here.

(32) Defeasible Modus Tollens:
Students are (normally) unemployed.
Dick is employed.

Dick is (presumably) not a student.

(33) Students are (normally) adults.
Adults are (normally) not students.
Dick is a student.

Dick is (presumably) an adult.

In the first example, if the first rule applies to the domain in which Dick is a student and he is employed, the set of normal worlds in that domain becomes the empty set because there is no possible world in which Dick is unemployed. So it cannot apply to the domain, but it can apply to a larger domain because there is no incompatible rule that keeps the rule from applying to the larger domain. When the rule applies to the domain in which Dick is employed, only possible worlds in which Dick is not a student conform to the rule. Hence it is expected that Dick is not a student. This is compared with the second example. When Dick is a student in the current factual knowledge state, the first rule is in that domain. The second rule does not apply to the domain in which Dick is a student and adult. For in that domain it would make the set of normal worlds the empty set. The rule does not apply to a larger domain either, because we have the first rule which is incompatible with the rule. So the first rule is the only one that works in the domain in which Dick is a student. Without the first rule, we would expect that Dick is not an adult.

There are two cases where a rule in a domain does not apply to a (smaller) domain. One is where a domain to which a rule would otherwise apply contains an incompatible rule. The rule is an exception rule to the default rule in question. Consider the example in (27) again. The rule that adults are employed does not apply to the domain for the third rule which is exceptional to the first rule. Even in this case, the second rule must apply to the domain for the third rule. This is guaranteed by the new condition because there is no rule in any domain which is incompatible with that rule. The other case is where there are two rules which are incompatible in some domain, but the domain for one rule is not a subset of the one for the other. The two rules are only partially incompatible, so we cannot say that one rule is exceptional to the other. This is illustrated in the following.

(34) Nixon Diamond:
Republicans are nonpacifists.
Quakers are pacifists.
Dick is a republican and a quaker.

---

*Dick is presumably a pacifist.
*Dick is presumably a nonpacifist.

Suppose that the first two rules are true in a possible world w. Neither of the rules cannot hold in the domain in which Dick is a republican and a quaker. So when the current factual knowledge state is that domain, we do not expect Dick to be a pacifist or a nonpacifist. This is guaranteed by the new condition .

Before proceeding with other discussion, I want to point out that specifying conditions with respect to rules does not mean that we need patterns on domains, as in Veltman's theory. Once we derive a frame, we only consider the most normal worlds in a domain d, which is expressed as $f(w)(d)$. We do not have to consider the ordering of possible worlds which are not the most normal in the domain. Despite this, I discuss rules here because the main concern of this paper is to show how an inference is made and it totally depends on the interactions between rules.

I have talked about the frame of one possible world. Making inference is a matter of information, so we must consider a set of possible worlds and the frames from them. What inference can be made is determined by what rules are included in the frame. Different possible worlds include different default rules. A rule in a domain may apply to a certain domain in the frame of a possible world while the same rule does not apply to the same domain in the frame of another possible world. More significant difficulty comes from the fact that the frame of a possible world contains all rules which are true in that possible world, but that in an information state we only know some of the rules. In order to make a correct inference from the current information state s, we only have to consider the rules which have been known in s, regardless of what rules are actually true in each possible world in s. What we also know is that there is at least one possible world in s in which only the rules known so far are true. There may be more than one such world, and they share the same frame. I will call it a minimal frame. If a new rule is known, the possible worlds which give rise to the minimal frame will be eliminated from s, because such possible worlds would not support the new rule.

The minimal frame is written as $fm(s) = Min(\{f(w) \mid w \in s\})$. The question is how to get $fm(s)$. The notion of minimum is not easy to capture in the semantics for default rules because it is non-monotonic. If we want to find a minimal frame in the current information state, we need to find some monotonic aspect of a frame. One such is the set of normal worlds in the domain of W. A minimal frame has more normal worlds in the domain of W. This is clear from the definition of normal worlds. To repeat the definition of normal world, a possible world is a normal world in a domain d if and only if it conforms to every rule which applies to every domain $d' \subseteq d$ to which the possible world belongs. From the definition here, $N_{f(w)}(W)$ is the set of possible worlds which are normal with respect to every default rule in w. This set will reduce if a new default rule is added. So this will help us get the minimal frame. But comparing $N_{f(w)}(W)$ is not sufficient. Consider the following two discourse segments.

(35) Adults are employed.
College students are not employed.

(36) Adults are employed.
College students are not employed.
Poor college students are employed.

Suppose that there are two frames for the two discourses and that each frame does not contain any other rule than the given. In this case, $N_{f(w)}(W)$ is the same for the two frames: the set of possible worlds in which adults are employed and adults are not college students. So $N_{f(w)}(W)$ does not distinguish the two frames. We need to compare some other domains which are smaller than W. Even in this case, there must be some salient domain in which the set of normal worlds can be compared. Otherwise we can not pick out a minimal frame.

I need to show that there *is* a minimal frame. Suppose that there are two frames. One frame has one more rule than the other. When the additional rule is not exceptional to any other rule, it applies to every (larger) domain by Condition 3 (original and revised), and reduces the sets of normal worlds in every (larger) domain including W. Then the comparison of $N_{f(w)}(W)$ will indicate which frame is a minimal one. When the additional rule is exceptional to a rule which is included in both frames, we cannot compare the set of normal worlds in the domain to which the additional rule is added. But still we can compare the set of normal worlds in W. This is illustrated in (35). Suppose that the first rule is the one shared by both frames, and that the second is the additional rule. In the frame without the additional rule, $N_{f(w)}(W)$ is the set of possible worlds in which adults are employed. In the frame with the additional rule, on the other hand, $N_{f(w)}(W)$ is the set of possible worlds in which adults are employed and are not college students. The additional rule reduces the set of normal worlds in W. When the additional rule is exceptional to a rule which is exceptional to another rule, the additional rule does not change $N_{f(w)}(W)$. This is illustrated in (36). Suppose that the first two rules are shared by both frames and the third is the additional rule. The comparison of $N_{f(w)}(W)$ does not tell us which one is minimal. Instead, we need to compare the set of normal worlds in the domain to which the second one is added. We do not need to compare other domains. When a rule is added to a domain, it reduces the set of normal worlds in the domain W or in a domain which contains an exception rule to which the additional rule is exceptional.

From the discussion so far, we can determine a minimal frame among $\{f(w) \mid w \in s\}$ as follows:

(37) A frame is minimal if

1. it has at least as few domains in which exception rules are as any other frame, and

2. it has as many normal worlds as any other frame in the domain W and in every domain which contains an exception rule to which another rule is exceptional.

When a frame is minimal, it does not have to have as many normal possible worlds as other frames in every domain. This is because the number of normal worlds in a domain is determined by the interactions between default rules, as the conditions on frames show. We only consider the domains specified above.

Now it is possible to interpret a sentence with the operator *presumably*.

Definition 19
For a set of possible worlds $s \subseteq W$, let $fm(s) = Min(\{f(w) \mid w \in s\})$.

$$s[\triangleright\phi] = \begin{cases} s \text{ if } N_{fm(s)}(s)[\phi] = N_{fm(s)}(s) \\ \emptyset \text{ otherwise.} \end{cases}$$

In the interpretation of the operator *presumably*, I use the set of normal worlds $N_{fm(s)}(s)$, not simply $fm(s)$. A sentence with the operator *normally* is interpreted with respect to a domain in which the antecedent clause of the rule holds. What is crucial is what is normally the case when the antecedent clause holds. We do not have to consider what (exception) rules hold in its smaller domains. Unlike a sentence with the operator *normally*, the operator *presumably* is interpreted with respect to the current information state, that is, considering all default rules that are already known. Therefore we have to use normal worlds which are defined with respect to a domain and every smaller domain. At this point, I cannot show that the notion of normal world is exactly what we need. The only thing we can do is show that by using the notion of normal world, we can explain every property related to the operator *presumably*. This will be done below.

5.3 Negation of a Default Rule

Veltman's theory has a problem with the interpretation of the negation of a default rule. The new theory does not have such a problem. I will prove this by showing that the negation of a default and a negative rule yield different results. Suppose that $\neg(\phi \succ \psi)$ is a sentence to be interpreted.

(38) $s[\neg(\phi \succ \psi)] = s \backslash s[(\phi \succ \psi)]$
$= s\backslash\{w \in s \mid f(w)(W[\phi])[\psi] = f(w)(W[\phi])\}$
$= \{w \in s \mid f(w)(W[\phi])[\psi] \neq f(w)(W[\phi])\}$

From this we get the set of possible worlds in which $(\phi \succ \psi)$ does not hold. Even when all possible worlds in which $(\phi \succ \psi)$ is true are eliminated, it is not guaranteed that $(\phi \succ \neg\psi)$ holds. This becomes clear from the interpretation of $(\phi \succ \neg\psi)$.

(39) $s[(\phi \succ \neg\psi)] = \{w \in s \mid f(w)(W[\phi])[\neg\psi] = f(w)(W[\phi])\}$
$= \{w \in s \mid f(w)(W[\phi])\backslash f(w)(W[\phi])[\psi]) = f(w)(W[\phi])\}$
$= \{w \in s \mid f(w)(W[\phi])[\psi] = \emptyset\}$

When an information state is updated with $\neg(\phi \succ \psi)$, some of the remaining worlds have a frame $f(w)$ such that $f(w)(W[\phi])[\psi]$ does not become the empty set. $(\phi \succ \neg\psi)$ is false in those worlds. The existence of such possible worlds is the evidence that $(\phi \succ \neg\psi)$ is stronger than $\neg(\phi \succ \psi)$. That is, the second sentence can be followed by the negation of the first sentence.

(40) $s[\neg(\phi \succ \psi) \wedge \neg(\phi \succ \neg\psi)] = (s\backslash s[(\phi \succ \psi)])[\neg(\phi \succ \neg\psi)]$
$= \{w \in s \mid f(w)(W[\phi])[\psi] \neq f(w)(W[\phi])\}\backslash\{w \in s \mid f(w)(W[\phi])[\psi]$
$\neq f(w)(W[\phi])\}[(\phi \succ \neg\psi)]$
$= \{w \in s \mid f(w)(W[\phi])[\psi] \neq f(w)(W[\phi]) \text{ and } f(w)(W[\phi])[\psi] \neq \emptyset\}$

The resulting possible worlds are those in which $\neg(\phi \succ \psi)$ is true and $(\phi \succ \neg\psi)$ is false. The set of the possible worlds does not have to be the empty set and it proves that $(\phi \succ \neg\psi)$ is stronger than $\neg(\phi \succ \psi)$. In the new theory of default rules, the interpretations of the two different default rules are not the same. This is what we expect from the correct theory of default rules.

5.4 Explanation of Inference Rules

At the beginning, we have seen various inference rules in relation to default rules. In this section I will show how the properties can be explained in the new theory.

(41) Defeasible Modus Ponens:
Lions have four legs.
Simba is a lion.

Simba presumably has four legs.

The sentence 'Lions have four legs' is understood as 'for every individual a, if a is a lion, it is normally the case that a has four legs'. This rule applies to every domain in which an individual is a lion. When we make an inference for an individual, we can ignore the universal quantifier. When we make an inference about Simba, the relevant domain is the one in which Simba is a lion ($= p$). The rule makes the worlds in which Simba has four legs ($= q$) at least as normal as any other world in the domain. Without any other rule known at the moment, the current information state is the domain in which Simba is a lion. Suppose that the sentence 'Simba is a lion' is p, and 'Simba has four legs' is q. $N_{fm(s)}(||p||) \subseteq ||q||$. Hence Simba presumably has four legs.

Now consider a case where the inference is canceled.

(42) Lions have four legs.
Simba is a lion. Simba has three legs.

---

*Simba presumably has four legs.

In this case, the current information state s is the set of possible worlds in which Simba is a lion and has three legs. For any possible world in s, if the rule 'Lions have four legs' applied to this domain, $f(w)(s)$ would become the empty set, which is prohibited by the condition that a frame be coherent. Therefore $f(w)(s)$ contains possible worlds in which Simba does not have four legs. Therefore assuming that 'Simba has four legs' is q, $N_{fm(s)}(s)[q]$ does not become $N_{fm(s)}(s)$. So we do not expect that Simba has four legs. A case of defeasible Modus Tollens and a case in which defeasible Modus Ponens wins defeasible Modus Tollens are already discussed sufficiently in a previous section.

Another characteristic of generic statements is that they allow exceptions. This example is good for the explanation why a set of normal worlds is used in the interpretation of *presumably*.

(43) Exception rule:
Adults are employed.
College students are not employed.
Dick is an adult.

Dick is presumably employed and not a college student.

The conditions on frame allow exceptions. So it is not necessary to show this. Even though we admit the first statement without excluding the second statement. In this case we cannot say that adults are normally not college students, because some normal adults can be college students. But when we interpret the operator *presumably*, we use a set of normal worlds in the domain of the current factual information state s. When the minimal frame is $fm(s)$, $fm(s)(s)$ contains some possible worlds in which Dick is a college student. But these worlds are not normal worlds because Dick is not a normal student in those worlds. So they are not included in $N_{fm(s)}(s)$. The set of normal worlds tells us that Dick is not absolutely normal in that he is not a normal student. This makes us expect that he is presumably not a college student.

Let's look at a case where two default rules are not completely incompatible.

(44) Nixon Diamond:
Republicans are nonpacifists.
Quakers are pacifists.
Dick is a quaker.
Dick is a Republican.

---

*Dick is presumably a pacifist.
*Dick is presumably a nonpacifist.

A rule in a larger domain applies to a smaller domain only when there is no rule which is incompatible with the rule and which, but for the rule, could apply to the smaller domain. The first two rules cannot apply to the domain in which Dick is a quaker and a Republican, as discussed in the previous discussion. Then $N_{fm(s)}(\{w \in W \mid \text{Dick is a quaker and a Republican in } w\})$ is the same as $\{w \in W \mid \text{Dick is a quaker and a Republican in } w\}$. Dick is a pacifist in some of these worlds, but he is a nonpacifist in the others. Therefore we cannot say either that Dick is presumably a pacifist or that Dick is presumably a nonpacifist.

Let's consider a case where two rules are in a domain, and only one of them is canceled by some factual information in a smaller domain.

(45) Graded Normality:
Birds normally fly
Birds normally have two legs.
Tweety is a bird.
Tweety does not fly.

Tweety presumably have two legs.

The first two rules are in the domain of $\{w \in W \mid \text{Tweety is a bird in } w\}$ and may apply to its smaller domains including the domain of $s = \{w \in W \mid \text{Tweety is a bird and does not fly in } w\}$, which is the current factual information state. Among the first two rules, the first one does not apply to the domain s, because it would make the set of normal worlds of s the empty set. Only the second rule applies to the domain s, and so without any other rule than the given rules, we predict that Tweety presumably have two legs.

6. Conclusion

Veltman claims that each domain in a frame must have a form of a pattern for the semantics of default rules. He uses the ordering of possible worlds based on a set of rules in a domain. This makes the theory not only complex but inappropriate for the negation of a default rule and other sentences which require the meaning of a default rule to be defined in a possible world. The problem with the interpretation of the negation of a default rule can be solved by assuming a more complex structure of a frame. But Veltman's theory has a more fundamental problem. By assuming that an information state has two components which have different structures, the theory will cause much more complexity in interpreting sentences in which a default rule is embedded.

A more fundamental solution is to define the meaning of a default rule as a set of possible worlds, just like other sentences. Then even when a default rule is embedded in a complex sentence, it is interpreted just like ordinary expressions embedded in complex sentences. It is the correct direction to solve Veltman's problems with the interpretation of default rules. But no satisfactory theory is proposed in this direction. The main reason is that when a default rule is interpreted with respect to a possible world, it is difficult to explain inferential properties in relation to default rules. The main difficulty comes from nonmonotonicity in default logic. In this paper, I have shown that despite the nonmonotonicity that is observed in making an inference, there is still some aspect which indicates which possible world includes the least number of rules in a frame. By introducing the notion of minimal possible world in an information state, we can make the right inferences based on the default rules that are already know, even if the information state changes.

One more difference from Veltman's theory is that our theory assumes a simple structure of a frame. A frame takes a set of possible worlds as an input, and yields its subset as the value, instead of a set of rules or the ordering between possible worlds. When we derive a frame $f(w)$ from the set of rules that are true in a possible world w, we need to know if each rule can apply to a certain domain. But once a frame is determined, we do not need to know what rules are in which domain. All we need to know is which possible world is the *most* normal in each domain. We do not need to know which possible world is more normal than which possible world if they are not the most normal worlds.

References

Asher, Nicholas and Michael Morreau. 1991. Commonsense entailment: A modal theory of nonmonotonic reasoning. In *Proceedings of the 12th IJCAI*, pages 387–392, San Mateo, Cal. Morgan Kaufmann.

Asher, Nicholas and Michael Morreau. 1995. What some generic sentences mean. In G.N. Calson et al. (eds.), editor, *The Generic Book*. University of Chicago Press, Chicago, pages 300–338.

Carlson, Gregory N. 1977. *Reference to Kinds in English*, volume Ph.D. dissertation. University of Massachesetts, Amherst, Published 1980 by Garland Press, New York.

Carlson, Gregory N. 1982. Generic terms and generic sentences. *Journal of Philosophical Logic*, 11:145–181.

Carlson, Gregory N. 1988. Truth conditions of generic sentences: Two contrasting views. In M. Krifka (ed.), editor, *Genericity in Natural Language*. SNS-Bericht 82-42, University of Tübingen, pages 31–51.

Delgrande, J. 1987. A semantics for defaults using conditional logic. *Artificial Intelligence*, 33:105–130.

Groeneveld, Willem and Frank Veltman. 199? Inference systems for update semantics. *Dyana Deliverable Report*, pages 229–273.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*, volume Ph.D dissertaton. University of Massachesetts, Amherst.

Kratzer, Angelika. 1981. The notional category of modality. In H.-J. Eikmeyer and H. Rieser (eds.), editors, *Words, Worlds, and Contexts: New Approaches to Word Semantics*. de Gruyter, Berlin, pages 38–74.

Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12:607–653.

Krifka, Manfred. 1987. *An Outline of Genericity*, volume partly in collaboration with Claudia Gerstner. SNS-Bericht 87-23, University of Tübingen.

Lewis, David. 1973. *Counterfactuals*. Basil Blackwell, Oxford.

Link, Godehard. 1995. Generic information and dependent generics. In G. N. Carlson et al. (eds.), editor, *The Generic Book*. University of Chicago Press, Chicago, pages 358–382.

Pelletier, F.J. and N. Asher. 1997. Generics and defaults. In Johan van Benthem and Alice ter Meulen (eds.), editors, *Handbook of Logic and Language*. Elsevier, Amsterdam, pages 1125–1177.

Stalnaker, Robert. 1987. A theory of conditionals. In N. Rescher (ed.), editor, *Studies in Logical Theory*. Blackwell, Oxford, pages 98–112.

Thomason, R.H. 1997. Nonmonotonicity in linguistics. In Johan van Benthem and Alice ter Meulen (eds.), editors, *Handbook of Logic and Language.* Elsevier, Amsterdam, pages 777–831.

Veltman, Frank. 1995. Defaults in update semantics. *Forthcoming in The Journal of Philosophical Logic.*

Submitted on: August 8, 2000
Accepted on: September 22, 2000