A Foundation for Metareasoning Part II: The Model Theory

GIOVANNI CRISCUOLO, *Dipartimento di Scienze Fisiche, University of Naples, 80125 Napoli, Italy.*
E-mail: vanni@na.infn.it

FAUSTO GIUNCHIGLIA, *Dipartimento di Informatica e Studi Aziendali, University of Trento, 38100 Trento, Italy, and ITC–IRST, Centro per la Ricerca Scientifica e Tecnologica, 38050 Trento, Italy.*
E-mail: fausto@itc.it

LUCIANO SERAFINI, *ITC–IRST, Centro per la Ricerca Scientifica e Tecnologica, 38050 Trento, Italy.*
E-mail: serafini@itc.it

Abstract

OM pairs are our proposed framework for the formalization of metareasoning. OM pairs allow us to generate deductively the object theory and/or the metatheory. This is done by imposing, via appropriate reflection rules, the relation we want to hold between the object theory and the metatheory. In a previous paper we have studied the proof theoretic properties of OM pairs. In this paper we study their model theoretic properties, in particular we study the relation between the models of the metatheory and the object theory; and how to use these results to refine the previous analysis.

Keywords: Metatheory, reflection, contextual reasoning.

1 Introduction

The *'Meta' property* is not a property which can be ascribed to a single theory; it is rather a relation between two theories. A theory is said to be *meta* of another, not necessarily distinct, theory if it is *about* this other theory. The latter theory is often said to be the *object* of the former theory. The metatheory speaks about the object theory by using a special set of predicates, called *metapredicates*, each of which is supposed to represent a property about the object theory. By property here we mean a set of formulas of the language of object theory. Examples of properties are: 'being true in a model of the object theory', 'being a theorem in the object theory', 'being a theorem in the object theory extended with the set of axioms X' etc.

A *reflection principle* [3] expresses the correctness of the metatheory when a metapredicate is interpreted in a certain property about the object theory. A reflection principle, therefore, is a statement of the form: *'if a metaformula, expressing a certain proposition about the object theory, is a theorem of the metatheory, then such a proposition is actually true'*. The definition of metatheories that satisfy a reflection principle has been one of the main research issues in the area of Formal Logic, Computer Science, and Artificial Intelligence. Such metathe-

ories are usually built by adding to a theory special formulas, or special inference rules, or combinations of the two for the metapredicates.

In a Part I [2] we introduced a new framework called OM pairs for a uniform specification of metatheories that satisfy different reflection principles. OM pairs allow us to specify metatheories via a set of inference rules, called *reflection rules*. Reflection rules are special inference rules between the object theory and the metatheory (which are supposed to be distinct theories). They allow one to infer formulas in the metatheory starting from formulas in the object theory and vice versa.

Once a metatheory is specified in the previous terms, it is necessary to characterize it, namely, to verify whether it satisfies a given reflection principle. This amounts to verifying, whether the metapredicate correctly represents the property expressed in the reflection principle. This characterization concerns two aspects: the set of theorems of the metatheory, and its class of models. The former problem has been studied in Part I [2]. In this paper we concentrate on the latter. We propose a model theoretic characterization of the metatheory in terms of

> *the class of properties about the object theory (represented as set of formulas) which are the interpretations of the metapredicate in a model of the metatheory.* (1.1)

Notice that we consider a class of properties rather than a single one. This is because, in general, there is more than one interpretation of the metapredicate that satisfies the metatheory. (1.1) can be rephrased more formally as follows:

> *the class of sets of formulas P in the language of the object theory, such that there is a model m_P of the metatheory that interprets the metapredicate in P.* (1.2)

The core of the paper contains a characterization of metatheories containing a single metapredicate $\bullet$, and specified via OM pairs. We show how metatheories which respect reflection principles about different properties of the object theory, such as 'the set of formulas satisfiable in a specific model of the object theory', 'the set of formulas that are provable in an extension of the object theory' etc. can be defined via different combinations of reflection rules.

The model-theoretic characterization allows us to refine the characterization of the different reflection rules given in Part I [2]. We indeed refine the partial order on the strength of reflection rules given in that paper, by proving that it is indeed a strict partial order. The consequence of this is that each combination of reflection rules represents different reflection principles.

The paper is structured as follows: in Section 2 we recall the basic definition of OM pair given in Part I [2]. In Section 3 we characterize the models of the metatheory. The main idea is to define them as sets of object level formulas. This allows us to study the relations between the models of the metatheory and the object theory and, therefore, to make precise statements like 'the metatheory speaks about the object theory' and 'the object theory is the model of the metatheory'. Section 4 refines the analysis described in Section 3 of [2] and gives a strict ordering characterization of the reflection principles formalizables by OM pairs, based on the strength of the derivability relations and the sets of theorems generated. In Section 5, we generalize the notion of duality, given in [2], to models, and we show that it preserves satisfiability.

2 OM pairs

In this section we briefly recall the main definition of OM pairs given in Part I.

We start from two arbitrary theories O and M, presented as axiomatic formal systems, i.e. $O = \langle L_O, \Omega_O, \Delta_O \rangle$ and $M = \langle L_M, \Omega_M, \Delta_M \rangle$. O is called the *object theory*, M is called the *metatheory*. L_O (L_M) is the *language* of O (M), Ω_O (Ω_M) is the *set of axioms* of O (M), and Δ_O (Δ_M) is the *deductive machinery* (set of inference rules) of O (M). We call *Meta-formulas*, the formulas of L_M. We suppose that Δ_O and Δ_M are set of natural deduction style inference rules, as defined in [6].

$\vdash_{\text{O}}$ and $\vdash_{\text{M}}$, are the derivability relations defined by O and M, respectively; $\mathrm{TH}(O)$ and $\mathrm{TH}(M)$ are the set of theorems of O and M respectively. We then consider new kinds of inter-theory inference rules, called *bridge rules* [4], whose premisses and conclusions belong to different languages. Thus we may have a bridge rule with premisses in L_O and conclusion in L_M; and, vice versa, a bridge rule with premisses in L_M and conclusion in L_O. Bridge rules extend deductively $\mathrm{TH}(O)$ and $\mathrm{TH}(M)$, that is, new object and metatheorems may be proved by applying bridge rules. The set of theorems of the metatheory may be extended whenever the set of theorems of the object theory is extended, and vice versa. We focus on the following bridge rules, and we call them *reflection rules*.

$$\frac{A}{\bullet(\text{“}A\text{”})}\ \mathrm{Rup} \qquad \frac{\neg A}{\neg \bullet(\text{“}A\text{”})}\ \mathrm{Rup}^{\mathrm{n}} \qquad \frac{\bullet(\text{“}A\text{”})}{A}\ \mathrm{Rdw} \qquad \frac{\neg \bullet(\text{“}A\text{”})}{\neg A}\ \mathrm{Rdw}^{\mathrm{n}}$$

For each of the reflection rule above, we consider two versions, the unrestricted version, which can be applied with no restriction and the restricted version (denoted by adding the index), which can be applied under the following condition:

RESTRICTIONS: Rules labelled with index r are applicable if the premiss does not depend on any assumptions in the same theory.

For the rest of the paper we make the following simplifying hypotheses: L_O and L_M are propositional and L_M is the *propositional metalanguage* of L_O.[1]

DEFINITION 2.1 (Propositional metalanguage)
Given a logical language L, its propositional metalanguage is the propositional language $\bullet(\text{“}L\text{”})$, whose set of atomic wffs is the set $\{\bullet(\text{“}A\text{”}) : A \in L\}$.

DEFINITION 2.2 (OM pair)
An *Object-Meta Pair* (OM pair) is a triple $\langle O, M, (\mathrm{RR}) \rangle$ where $O = \langle L_O, \Omega_O, \Delta_O \rangle$, $M = \langle L_M, \Omega_M, \Delta_M \rangle$, L_M contains the propositional metalanguage of L_O, and (RR) is a set of reflection rules.

Given an object theory O, a metatheory M, and a set of reflection rules (RR), we say that $\mathrm{OM} = \langle O, M, (\mathrm{RR}) \rangle$ is the OM pair *composed of* O *and* M *connected by* (RR). Notationally, when it contains more than one bridge rule we represent (RR) by listing its elements separated by a $+$. One example of (RR) is $\mathrm{Rup} + \mathrm{Rdw}^{\mathrm{n}}$.

A wff $A \in L_O \cup L_M$ is derivable from a set of assumptions $\Gamma \subseteq L_O \cup L_M$, in OM, in symbols $\Gamma \vdash_{\text{OM}} A$, if there is a deduction in OM of A from Γ that applies Δ_O, Δ_M, the axioms of O and M, and the reflection rules of OM. A is *provable in* (*a theorem of*) OM, abbreviated as $\vdash_{\text{OM}} A$, if it is derivable from the empty set.

$\mathrm{TH}_{\mathrm{OM}}(O) \subseteq L_O$ is the set of formulas of L_O provable in OM. $\mathrm{TH}_{\mathrm{OM}}(M) \subseteq L_M$ is the set of formulas of L_M provable in OM. Trivially $\mathrm{TH}(O) \subseteq \mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}(M) \subseteq$

[1] In Part I [2] we have shown how the results in the propositional case can be generalized to the first-order case.

$\mathrm{TH}_{\mathrm{OM}}(M)$. Given an OM pair OM $= \langle O, M, (\mathrm{RR})\rangle$ we use the term *object theory of* OM *(generated by* (RR)*)* to denote any theory $O' = \langle L_O, \Omega'_O, \Delta_O\rangle$ whose set of theorems is $\mathrm{TH}_{\mathrm{OM}}(O)$ and the term *metatheory of* OM *(generated by* (RR)*)* to denote any theory $M' = \langle L_M, \Omega'_M, \Delta_M\rangle$ whose set of theorems is $\mathrm{TH}_{\mathrm{OM}}(M)$. Trivially TH(*O'*)= $\mathrm{TH}_{\mathrm{OM}}(O)$ and TH(*M'*)=$\mathrm{TH}_{\mathrm{OM}}(M)$.

We consider the OM pairs with the following combinations of reflection rules:

1. $\emptyset$
2. $\mathrm{Rdw}_{\mathrm{r}}$
3. $\mathrm{Rup}_{\mathrm{r}}$
4. Rdw
5. Rup
6. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$
7. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
8. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
9. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$
10. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
11. $\mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$
12. $\mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
13. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$
14. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$
15. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
16. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$
17. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$
18. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$
19. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}^{\mathrm{n}}$
20. $\mathrm{Rup} + \mathrm{Rdw}$
21. $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$
22. $\mathrm{Rup} + \mathrm{Rdw}^{\mathrm{n}}$

3 A model-theoretic characterization

The goal of this section is to understand the relation between the models of the metatheory and the object theory. For our purpose it is enough to consider the case in which L_M is the propositional metalanguage of L_O, i.e. all its atomic wffs are of the form $\bullet(\text{“}A\text{”})$ for some wff A of L_O. Being a propositional theory, the models of M are truth assignments for the atomic wffs of L_M. Any interpretation m of L_M identifies a subset $S_m \subseteq L_O$ according to the following definition:

$$S_m = \{A \in L_O \mid m(\bullet(\text{“}A\text{”})) = \textit{True}\}.$$

We can therefore consider the set of interpretations of L_M to be the powerset of the wffs in L_O.

For any set (RR) of reflection rules, we characterize:

- the models of the metatheory generated by (RR) in terms of O;
- the set of theorems of the object theory generated by (RR) in terms of the models of M.

The first result allows us to make precise in which sense *the metatheory speaks about the object theory*. The second result allows us to make precise the sense in which *the object theory is a model of the metatheory*.

We use the following notation. Let m be an interpretation of L_M, we say that m is a model of Γ, in symbols $m \models \Gamma$, to mean that m satisfies all wffs in Γ (according to the classical definition of satisfiability).

Furthermore we say that a set of wffs $\Gamma \subseteq L$ is a *classical propositional theory* if it contains all the classical tautologies and it is closed under modus ponens. Γ is consistent if $\bot \notin \Gamma$. Γ is *maximal* if for any wff $A \in L$, $A \in \Gamma$ or $\neg A \in \Gamma$. We also use the following notation. For any theory $T = \langle L, \Omega, \Delta\rangle$ and any $\Gamma \subseteq L$, $T + \Gamma$ denotes the extended theory $\langle L, \Omega \cup \Gamma, \Delta\rangle$. For any set of wffs $\Gamma \subseteq L_O$, $\mathrm{Rup}(\Gamma)$ denotes the set of wffs $\{\bullet(\text{“}A\text{”}) : A \in \Gamma\} \subseteq L_M$. We sometimes use $\bullet(\text{“}\Gamma\text{”})$ instead of $\mathrm{Rup}(\Gamma)$. Analogously for any set of wffs $\Gamma \subseteq L_M$, $\mathrm{Rdw}(\Gamma)$ denotes the set of wffs $\{A : \bullet(\text{“}A\text{”}) \in \Gamma\} \subseteq L_O$.

3.1 The models of the metatheory

Let us start by considering the metatheories generated by combinations of reflection rules which are characterizable axiomatically, i.e. the metatheories for which $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \Phi)$, with Φ being some schematic metaformula. ([2, Section 4] gives an in depth analysis of the axiomatic characterizability of metatheories.) In this case the (obvious) idea is to look for the subclass of interpretations of L_M which satisfies Ω_M and Φ.

THEOREM 3.1
Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR);

1. if $(\mathrm{RR}) =_M \mathrm{Rdw_r}$, then m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$, if and only if it is a model of Ω_M;
2. if $(\mathrm{RR}) =_M \mathrm{Rup_r}$, then m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$, if and only if it is a model of Ω_M, and $\mathrm{TH}(O) \subseteq m$.

Item (i) of Theorem 3.1 states that any model of M is a model of the metatheory of OM and vice versa. m is a model of the metatheory of OM independently of the object theory. Item (ii) states that the models of the metatheory of OM are all the models of Ω_M which contain the theorems of the object theory.

PROOF. [Theorem 3.1] These results are a direct consequence of [2, Theorem 4.1]. This theorem states the (obvious) result that reflection down leaves the metatheory unchanged, and we have $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M)$, while reflection up adds to the metatheory all the theorems of the form $\bullet(\text{“}A\text{”})$, with $A \in \mathrm{TH}(O) = \mathrm{TH}_{\mathrm{OM}}(O)$. ∎

THEOREM 3.2
Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR). An interpretation m of L_M is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$ if and only if, $m \models \Omega_M$ and:

1. If (RR) is $\mathrm{Rup_r} + \mathrm{Rdw}$ then, $\mathrm{TH}(O) \subseteq m$ and m is a *classical propositional theory*.
2. If (RR) is $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$, then $\mathrm{TH}(O) \subseteq m$ and m is a *consistent classical propositional theory*.
3. If (RR) is $\mathrm{Rup} + \mathrm{Rdw}$, then $\mathrm{TH}(O) \subseteq m$ and m is a *maximal classical propositional theory*.
4. If (RR) is $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup_r^n}$, then $\mathrm{TH}(O) \subseteq m$ and m is a *consistent and maximal classical propositional theory*.
5. If (RR) is either $\mathrm{Rup_r^n} + \mathrm{Rdw}$ or $\mathrm{Rup^n} + \mathrm{Rdw}$, then either $O + m$ is *consistent*, or $m = \emptyset$.

To understand Theorem 3.2, let us consider the case with $\Omega_M = \emptyset$. In this situation the models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are completely determined by O and (RR). (These observations can be easily generalized. In fact, from Theorem 3.2, $\Omega_M \neq \emptyset$ restricts us to the models of Ω_M.) Consider item (1). The models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are the sets of theorems of the extensions of the object theory, i.e. all the theories which can be obtained by adding a set of axioms Γ to O. Therefore, if A is a logical consequence of a set of formulas Γ, then $\bullet(\text{“}A\text{”})$ is also a logical consequence of $\bullet(\text{“}\Gamma\text{”})$. Consider item (2). The models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are the sets of theorems of the consistent extensions of the object theory. Notice that, as there are no consistent extensions of an inconsistent theory, if O is inconsistent then $\mathrm{TH}_{\mathrm{OM}}(M)$ is inconsistent as well. Consider item (3). The models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are the maximal (possibly inconsistent) extensions of the object theory. Consider item (4). The models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are the

maximal consistent extensions of the object theory. Since the set of maximal consistent sets containing the object theory is isomorphic to the set of models of the object theory, $\bullet(\text{“}A\text{”})$ can be interpreted as 'A is true in m'. Finally, consider item (5). The models of $\mathrm{TH}_{\mathrm{OM}}(M)$ are the sets of formulas which can be added to the object theory without making it inconsistent. If, for instance, A is derivable from Γ in O, the object theory extended by Γ and $\neg A$ is inconsistent. This corresponds to the fact that in any model m of $\mathrm{TH}_{\mathrm{OM}}(M)$ it is never the case that Γ and $\neg A$ both belong to m. A further example: if the object theory is inconsistent (i.e. all formulas are provable) then it cannot be consistently extended. In this case the only model of the metatheory of OM is the empty set.

PROOF. [Theorem 3.2] This proof is based on the result of Theorems 4.2 and 4.3 in [2]. In these theorems the metatheory of OM is characterized in terms of the following formulas:

$$\bullet(\text{“}A \supset B\text{”}) \supset (\bullet(\text{“}A\text{”}) \supset \bullet(\text{“}B\text{”})) \qquad \text{(K)}$$

$$\neg \bullet (\text{“}A\text{”}) \supset \bullet(\text{“}\neg A\text{”}) \qquad \text{(Comp)}$$

$$\neg \bullet (\text{“}\bot\text{”}) \qquad \text{(nTbot)}$$

$$\{\bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_k\text{”}) \supset \neg \bullet (\text{“}A\text{”}) :\vdash_{O+A_1,\ldots,A_n} \neg A\} \qquad \text{(Cons)}$$

where A, B are understood as schematic variables (parameters) ranging over L_O.

Item (1). In this case $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O)) + (K))$, as proved by [2, Theorem 4.2]. Then, $m \models \mathrm{TH}_{\mathrm{OM}}(M)$ if and only if $m \models \Omega_M$, $m \models \mathrm{Rup}(\mathrm{TH}(O))$ and $m \models (\mathrm{K})$. Suppose that m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$, then $m \models \Omega_M$ and $\mathrm{TH}(O) \subseteq m$. m is a classical propositional theory as it contains all the propositional tautologies (contained in $\mathrm{TH}(O)$), and it is closed under *modus ponens* since $m \models (\mathrm{K})$. Vice versa, suppose that $m \models \Omega_M$ and m is a classical propositional theory which contains $\mathrm{TH}(O)$, then $m \models \Omega_M$, by hypothesis, $m \models \mathrm{Rup}(\mathrm{TH}(O))$ as $\mathrm{TH}(O) \subseteq m$, and $m \models (\mathrm{K})$ as m is closed under logical consequence.

Item (2). In this case $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O)) + (K) + (\mathrm{nTbot}))$, as proved by [2, Theorem 4.2]. Then, $m \models \mathrm{TH}_{\mathrm{OM}}(M)$ if and only if, $m \models \Omega_M$, $m \models \mathrm{Rup}(\mathrm{TH}(O))$, $m \models (\mathrm{K})$ and $m \models (\mathrm{nTbot})$. From item *(1)*, if $m \models \mathrm{TH}_{\mathrm{OM}}(M)$, then $m \models \Omega_M$, and m is a classical propositional theory that contains $\mathrm{TH}(O)$. Furthermore $m \models (\mathrm{nTbot})$ implies that $\bot \notin m$ and therefore that m is consistent. Vice versa, if $m \models \Omega_M$ and m is a consistent classical propositional theory which contains $\mathrm{TH}(O)$ then, by item *(1)*, $m \models \mathrm{Rup}(\mathrm{TH}(O))$ and $m \models (\mathrm{K})$. The consistency of m implies that $\bot \notin m$ and therefore that $m \models (\mathrm{nTbot})$.

Item (3). In this case $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O)) + (K) + (\mathrm{Comp}))$, as proved by [2, Theorem 4.2]. Then, $m \models \mathrm{TH}_{\mathrm{OM}}(M)$ if and only if $m \models \Omega_M$, $m \models \mathrm{Rup}(\mathrm{TH}(O))$, $m \models (\mathrm{K})$ and $m \models (\mathrm{Comp})$. From item *(1)* of this theorem, if $m \models \mathrm{TH}_{\mathrm{OM}}(M)$, then $m \models \Omega_M$, and m is a classical propositional theory that contains $\mathrm{TH}(O)$. Furthermore $m \models (\mathrm{Comp})$ implies that, for any object wff A, $A \in m$ or $\neg A \in m$, i.e. m is maximal. Vice versa, if $m \models \Omega_M$ and m is a maximal classical propositional theory which contains $\mathrm{TH}(O)$, then, by item (1) of this theorem, $m \models \mathrm{Rup}(\mathrm{TH}(O))$ and $m \models (\mathrm{K})$. Furthermore m being maximal implies that $A \in m$ or $\neg A \in m$ and therefore that $m \models (\mathrm{Comp})$.

Item (4). In this case $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O)) + (K) + (\mathrm{Comp}) + (\mathrm{nTbot}))$, as proved by [2, Theorem 4.2]. Item (4) is therefore a consequence of items (2) and (3) of this theorem.

Item (5). In this case $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + (\mathrm{Cons}))$, as proved by [2, Theorem 4.3]. Let us start with the only if direction. Let m be a model of $\mathrm{TH}_{\mathrm{OM}}(M)$. Then, m is a model

of Ω_M, and by [2, Theorem 4.3], for any $A_1, \ldots, A_n, A$ such that $\vdash_{O+A_1,\ldots,A_n} \neg A$

$$\text{if } A_1, \ldots, A_n \in m, \text{ then } A \notin m.$$

Suppose that $O + m$ is inconsistent. Then, for any formula A, $\vdash_{O+m} \neg A$, this implies that there is a finite number of wffs $A_1, \ldots, A_n \in m$ such that $\vdash_{O+A_1,\ldots,A_n} \neg A$. This implies that $A \notin m$. As this is true for any formula A, m must be the empty set.

As far as the only if direction, to prove that $m \models \mathrm{TH}_{\mathrm{OM}}(M)$ we exploit again [2, Theorem 4.3] and prove that

$$m \models M + \{\bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”}) \supset \neg \bullet(\text{“}A\text{”}) : \vdash_{O+A_1,\ldots,A_n} \neg A\}. \tag{3.1}$$

If $m = \emptyset$ then $m \models \neg \bullet(\text{“}A\text{”})$ for any formula A, and since m satisfies Ω_M, then (3.1) is true. Suppose instead $m \neq \emptyset$ and $m \models \Omega_M$. Then, suppose that $m \models \bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”})$. If $\neg A$ is derivable in $O + A_1, \ldots, A_n$, then A cannot be in m, otherwise $O + m$ would be inconsistent. This implies $m \not\models \bullet(\text{“}A\text{”})$, that is $m \models \neg \bullet(\text{“}A\text{”})$, which implies (3.1). ∎

Let us now consider the combinations of reflection rules which are not characterizable axiomatically. We start with $\mathrm{Rup_r} + \mathrm{Rdw_r}$. For any set x of interpretations of L_M we define:

$$\begin{aligned} \bigcap x &= \{A \in L_O : \text{for all } m \in x,\ A \in m\} \\ \bigcup x &= \{A \in L_O : \text{there exists an } m \in x\ A \in m\}. \end{aligned}$$

THEOREM 3.3
Let OM be an OM pair composed of O and M connected by $\mathrm{Rup_r} + \mathrm{Rdw_r}$. The set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$ is the largest subset of interpretations of L_M that satisfies (3.2) in x:

$$x = \{m \models \Omega_M : \mathrm{TH}(O + \bigcap x) \subseteq m\}. \tag{3.2}$$

Theorem 3.3 states that m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$ iff it satisfies Ω_M and contains the set of formulas which are logical consequences of the intersection of all the models of $\mathrm{TH}_{\mathrm{OM}}(M)$. This theorem does not provide an explicit definition of the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$.[2] Indeed, to prove that a certain interpretation m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$ by exploiting Theorem 3.3, one needs, first to prove that a certain set $\mathfrak{M}$ of interpretations is contained in the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$ (by showing that $\mathfrak{M}$ is a solution of (3.2)), and then to prove that $m \in \mathfrak{M}$. Let us see some examples.

EXAMPLE 3.4
If Ω_M is the empty set, then a set $\mathfrak{M}$ which satisfies (3.2), is the set of $m \subseteq L_O$ which contains $\mathrm{TH}(O)$. Indeed any $m \in \mathfrak{M}$ satisfies an empty set of axioms Ω_M, and $\mathrm{TH}(O + \bigcap \mathfrak{M}) = \mathrm{TH}(O) \subseteq m$, because $\bigcap \mathfrak{M} = \mathrm{TH}(O)$.

EXAMPLE 3.5
If Ω_M contains the single axiom $\bullet(\text{“}p\text{”})$, then a set $\mathfrak{M}$ that satisfies (3.2) is the set of $m \subseteq L_O$ which contains $\mathrm{TH}(O + p)$.

EXAMPLE 3.6
If Ω_M contains the single axiom $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}q\text{”})$, then a set $\mathfrak{M}$ that satisfies (3.2) is the set of $m \subseteq L_O$ such that $\mathrm{TH}(O) \cup p \subseteq m$ or $\mathrm{TH}(O) \cup q \subseteq m$.

[2] This is the model theoretic counterpart of the fixpoint characterization of the metatheory generated by $\mathrm{Rup_r} + \mathrm{Rdw_r}$.

EXAMPLE 3.7
If Ω_M is the axiom $\neg \bullet$ ("p"), then a set $\mathfrak{M}$ that satisfies (3.2) is the set of subsets of L_O which contain $\mathrm{TH}(O)$ and does not contain p.

In the previous examples, notice that the models of $\mathrm{TH}_{\mathrm{OM}}(M)$ might not be closed under logical consequence. To force this closure we have to relax the restriction on the application of reflection up.

PROOF. [Theorem 3.3] The proof is in two steps. In the first step we prove that the set $\mathfrak{M}$ of models of the metatheory generated by $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$ satisfies (3.2). In the second step we prove that any other solution of (3.2) is contained in $\mathfrak{M}$.

First step
From [2, Theorem 4.4] we have that $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ are the smallest sets of wffs which satisfy the following equations:

$$x_O = \mathrm{TH}(O + \mathrm{Rdw}(x_M)) \tag{3.3}$$
$$x_M = \mathrm{TH}(M + \mathrm{Rup}(x_O)). \tag{3.4}$$

We have therefore that

$$\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M))))).$$

Let $\mathfrak{M}$ be the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$. $m \in \mathfrak{M}$ if and only if it satisfies Ω_M and satisfies all the formulas in $\mathrm{Rup}(\mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M))))$. We have therefore that $\mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M)) \subseteq m$, and that

$$\mathfrak{M} = \{m \models \Omega_M : \mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M))) \subseteq m\}.$$

Since $\bullet$("A") belongs to $\mathrm{TH}_{\mathrm{OM}}(M)$ if and only if for all $m \in \mathfrak{M}$, $A \in m$, we can conclude that $\mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M)) = \bigcap \mathfrak{M}$. We have therefore that:

$$\mathfrak{M} = \{m \models \Omega_M : \mathrm{TH}(O + \bigcap \mathfrak{M}) \subseteq m\}.$$

Second step
Let us prove that the $\mathfrak{M}$ is the largest set of interpretations of L_M which satisfies (3.2), namely, that if $\mathfrak{N}$ is a solution of (3.2), then $\mathfrak{N} \subseteq \mathfrak{M}$. To do this let us consider the following sequence of theories $M_0, M_1, \ldots$ (originally defined in the proof of [2, Theorem 4.4]):

$$M_0 = M \tag{3.5}$$
$$O_0 = O \tag{3.6}$$
$$M_{i+1} = M_i + \mathrm{Rup}(\mathrm{TH}(O_i)) \tag{3.7}$$
$$O_{i+1} = O_i + \mathrm{Rdw}(\mathrm{TH}(M_i)). \tag{3.8}$$

For any $i \geq 0$, let $\mathfrak{M}_i$ be the set of models of M_i. We prove by induction on i that for all $i \geq 0$, $\mathfrak{N} \subseteq \mathfrak{M}_i$.
Base case
$\mathfrak{N} \subseteq \{m : m \models \Omega_M\} = \mathfrak{M}_0$.

Step case
Since $\mathfrak{N}$ satisfies (3.2),

$$\mathfrak{N} = \{m \models \Omega_M : \mathrm{TH}(O + \bigcap \mathfrak{N}) \subseteq m\}.$$

The induction hypothesis $\mathfrak{N} \subseteq \mathfrak{M}_i$ implies that $\bigcap \mathfrak{M}_i \subseteq \bigcap \mathfrak{N}$, and therefore that

$$\mathfrak{N} \subseteq \{m \models \Omega_M : \mathrm{TH}(O + \bigcap \mathfrak{M}_i) \subseteq m\}.$$

Since $\mathfrak{M}_i$ is the set of models of M_i, we have that $\bigcap \mathfrak{M}_i = \mathrm{Rdw}(\mathrm{TH}(M_i))$. If we substitute $\bigcap \mathfrak{M}_i$ with $\mathrm{Rdw}(\mathrm{TH}(M_i))$ in the previous equation we obtain:

$$\mathfrak{N} \subseteq \{m \models \Omega_M : \mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}(M_i))) \subseteq m\}.$$

Notice that any interpretation m that satisfies Ω_M and $\mathrm{Rup}(O + \mathrm{Rdw}(\mathrm{TH}(M_i)))$ is a model of M_{i+1}, and therefore, we have that

$$\mathfrak{N} \subseteq \mathfrak{M}_{i+1}.$$

Thus we have shown that for any $i \geq 0$, $\mathfrak{N} \subseteq \mathfrak{M}_i$. This implies that:

$$\mathfrak{N} \subseteq \bigcap_{i \geq 0} \mathfrak{M}_i.$$

Since $\mathrm{TH}_{\mathrm{OM}}(M) = \bigcup_{i \geq 0} M_i$, the set of models of M is the intersection of the models of each M_i, i.e. $\mathfrak{M} = \bigcap_{i \geq 0} \mathfrak{M}_i$. Thus we conclude that $\mathfrak{N} \subseteq \mathfrak{M}$. ∎

The set of models of the metatheories generated by $\mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}} + \mathrm{Rdw}_{\mathrm{r}}$ and $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}_{\mathrm{r}}^{\mathrm{n}}$ can be characterized similarly to $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$. For a set of wff Γ, let $\neg\Gamma$ be defined as $\neg\Gamma = \{\neg A : A \in \Gamma\}$.

THEOREM 3.8
Let OM be an OM pair composed of O, M connected by $\mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}} + \mathrm{Rdw}_{\mathrm{r}}$. The set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$ is the largest subset of interpretations of L_M which satisfies (3.9) in x:

$$x = \{m \models \Omega_M : \mathrm{TH}(O + \bigcap x) \cap \neg m \neq \emptyset\}. \tag{3.9}$$

THEOREM 3.9
Let OM be an OM pair composed of O, M connected by $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}_{\mathrm{r}}^{\mathrm{n}}$. The set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$ is the largest subset of interpretations of L_M which satisfies (3.10) in x:

$$x = \left\{ \begin{array}{ll} m \models \Omega_M : & m \text{ is the set of theorems of an extension of } O \\ & \text{If } A \notin \bigcup x, \text{ then } \neg A \in m \end{array} \right\} \tag{3.10}$$

The proofs of these two theorems are analogous to that of Theorem 3.3.

$\mathrm{Rdw}_{\mathrm{r}}^{\mathrm{n}}$ implements a weak form of negation as failure. The models of the metatheory generated by $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}_{\mathrm{r}}^{\mathrm{n}}$ indeed are the extensions of O, which satisfy Ω_M and contain $\neg A$ whenever A is not provable in any extension of O.

3.2 The object theory in terms of the models of M

Our goal here is to understand the meaning of $\bullet$ in terms of properties of the object theory, i.e. which is the property of the object theory represented by $\bullet$. This allows us to make precise the sense in which the object theory can be considered a model of the metatheory.

We start from the weakest combinations of reflection rules and then move to those of increasing strength.

THEOREM 3.10
Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR). Let $\mathfrak{M}$ be the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$. Then:

1. If $(\mathrm{RR}) =_O \mathrm{Rdw}_{\mathrm{r}}$, then $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \bigcap \mathfrak{M})$.
2. If $(\mathrm{RR}) =_O \mathrm{Rup}_{\mathrm{r}}$, then $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O)$.

The underlying intuitions are the same as those described above for Theorem 3.1.

THEOREM 3.11
Let OM be an OM pair composed of O and M connected by a set of reflection rules (RR). Let $\mathfrak{M}$ be the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$. Then:

1. If $\mathrm{Rdw}_{\mathrm{r}} \leq_O (\mathrm{RR})$, then $\bigcap \mathfrak{M} \subseteq \mathrm{TH}_{\mathrm{OM}}(O)$.
2. If $\mathrm{Rup}_{\mathrm{r}} \leq_M (\mathrm{RR})$, then $\mathrm{TH}_{\mathrm{OM}}(O) \subseteq \bigcap \mathfrak{M}$.
3. if $\mathrm{Rdw}_{\mathrm{r}} \leq_O (\mathrm{RR})$ and $\mathrm{Rup}_{\mathrm{r}} \leq_M (\mathrm{RR})$, then $\mathrm{TH}_{\mathrm{OM}}(O) = \bigcap \mathfrak{M}$.

PROOF. [Theorem 3.11] Item (1) is a straightforward consequence of item (1) of Theorem 3.10. Item (2) is provable as follows. Suppose that $A \in \mathrm{TH}_{\mathrm{OM}}(O)$, then, by $\mathrm{Rup}_{\mathrm{r}}$, $\bullet(\text{“}A\text{”}) \in \mathrm{TH}_{\mathrm{OM}}(M)$ and therefore, $A \in \bigcap \mathfrak{M}$. Item (3) is the conjunction of (1) and (2). ∎

If the object theory is generated by a set of reflection rules weaker than $\mathrm{Rup}_{\mathrm{r}}$ or $\mathrm{Rdw}_{\mathrm{r}}$ (in the sense of Theorem 3.11), then the set of object theorems is not completely determined by the metatheory. Vice versa, from Theorem 3.11 we can conclude that the object theory (more precisely, its set of theorems) generated by a set of reflection rules which are at least as strong as $\mathrm{Rup}_{\mathrm{r}}$ and $\mathrm{Rdw}_{\mathrm{r}}$, is completely determined by the set of models (theorems) of the metatheory. Finally notice that in general, according to Theorem 3.11, an interpretation m of L_M with $m = \mathrm{TH}_{\mathrm{OM}}(O)$ is not necessarily a model of $\mathrm{TH}_{\mathrm{OM}}(M)$. Consider the following example.

EXAMPLE 3.12
Let OM be composed of the empty object theory and the metatheory with a set of axioms $\Omega_M = \{\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}q\text{”})\}$, connected by $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$. A model of the metatheory of OM contains either p or q. In particular $m_1 = \mathrm{TH}(O + p)$ and $m_2 = \mathrm{TH}(O + q)$ are both models of the metatheory of OM. On the other hand the intersection of these models (which is $\mathrm{TH}_{\mathrm{OM}}(O)$) does not contain p, nor q. Therefore, $\mathrm{TH}_{\mathrm{OM}}(O)$ is not a model of $\mathrm{TH}_{\mathrm{OM}}(M)$.

Theorem 3.11 provides the characterization of the object theory for all the combinations of reflection rules with the exception of $\mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}} + \mathrm{Rdw}$ and $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$. For these two special cases we have the following result.

THEOREM 3.13
Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR). If $\mathfrak{M}$ is the set of models of $\mathrm{TH}_{\mathrm{OM}}(M)$, then:

1. If (RR) is $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}$, then $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \bigcap \mathfrak{M})$;
2. If (RR) is $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$, then $\mathrm{TH}_{\mathrm{OM}}(O) = \bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m)$.

PROOF. [Theorem 3.13] Item (1). From [2, Theorem 4.3] we have

$$\mathrm{TH}_{\mathrm{OM}}(O) \quad = \quad \mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M))). \tag{3.11}$$

Since $\mathrm{Rdw}(\mathrm{TH}_{\mathrm{OM}}(M)) = \bigcap \mathfrak{M}$, (3.11) implies that $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \bigcap \mathfrak{M})$. To prove item (2), by [2, Theorem 4.3], we have

$$\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \{\textstyle\bigvee_{i=1}^{n} A_i : \bigvee_{i=1}^{n} \bullet(\text{“}A_i\text{”}) \in \mathrm{TH}_{\mathrm{OM}}(M)). \tag{3.12}$$

To prove that $\mathrm{TH}_{\mathrm{OM}}(O) \subseteq \bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O+m)$, we show that for any extra axiom $A_1 \vee \ldots \vee A_n$ added to O to obtain $\mathrm{TH}_{\mathrm{OM}}(O)$, we have that $A_1 \vee \ldots \vee A_n \in \bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m)$. For each $m \in \mathfrak{M}$, if $\bullet(\text{“}A_1\text{”}) \vee \ldots \vee \bullet(\text{“}A_n\text{”}) \in \mathrm{TH}_{\mathrm{OM}}(M)$ then $m \models \bullet(\text{“}A_1\text{”}) \vee \ldots \vee \bullet(\text{“}A_n\text{”})$, which implies that there is an $1 \leq i \leq n$ such that $A_i \in m$, and $A_1 \vee \ldots \vee A_n \in \mathrm{TH}(O+m)$. Therefore, for each $m \in \mathfrak{M}$, $\mathrm{TH}_{\mathrm{OM}}(O) \subseteq \mathrm{TH}(O + m)$; and therefore, $\mathrm{TH}_{\mathrm{OM}}(O) \subseteq \bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m)$.

Vice versa, let us prove $\bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m) \subseteq \mathrm{TH}_{\mathrm{OM}}(O)$. Let $A \in \bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m)$, and for each $m \in \mathfrak{M}$ let Π_m be a proof of A in $O + m$. Let M' be the following theory:

$$M + \mathrm{TH}_{\mathrm{OM}}(M) + \left\{ \neg(\wedge \bullet (\text{“}\Gamma_m\text{”})) : \begin{array}{l} m \in \mathfrak{M}, \text{ and } \Gamma_m \subseteq m \text{ is the finite (possibly} \\ \text{empty) set of formulas occurring as axioms in} \\ \Pi_m \end{array} \right\}$$

where $\wedge \bullet (\text{“}\Gamma_m\text{”})$ is the conjunction of the formulas in $\bullet(\text{“}\Gamma_m\text{”})$. Let us prove that M' is unsatisfiable. By contradiction let m be a model of M'. m is a model of $\mathrm{TH}_{\mathrm{OM}}(M)$ and therefore, $\Gamma_m \subseteq m$. This implies that $m \models (\wedge \bullet (\text{“}\Gamma_m\text{”}))$ and therefore that $m \not\models M'$. M' unsatisfiable implies that M' is finitely unsatisfiable, i.e. there is a finite sequence of sets $\Gamma_{m_1}, \ldots, \Gamma_{m_n}$ such $\mathrm{TH}_{\mathrm{OM}}(M) + \{\neg(\wedge \bullet (\text{“}\Gamma_{m_1}\text{”})), \ldots, \neg(\wedge \bullet (\text{“}\Gamma_{m_n}\text{”}))\}$ is unsatisfiable. This implies that

$$(\wedge \bullet (\text{“}\Gamma_{m_1}\text{”})) \vee \ldots \vee (\wedge \bullet (\text{“}\Gamma_{m_n}\text{”})) \in \mathrm{TH}_{\mathrm{OM}}(M)$$

and therefore that for each $\gamma_1, \ldots, \gamma_n$, with $\gamma_i \in \Gamma_{m_i}$,

$$\bullet(\text{“}\gamma_1\text{”}) \vee \ldots \vee \bullet(\text{“}\gamma_n\text{”}) \in \mathrm{TH}_{\mathrm{OM}}(M).$$

We have therefore that

$$\gamma_1 \vee \ldots \vee \gamma_n \in \mathrm{TH}_{\mathrm{OM}}(O).$$

A possible proof of A in $\mathrm{TH}_{\mathrm{OM}}(O)$ is therefore the following:

$$\dfrac{\dfrac{\begin{array}{c}\vdots \\ \bigwedge_{\gamma_i \in \Gamma_i}(\gamma_1 \vee \ldots \vee \gamma_n)\end{array} \quad \begin{array}{c}[\wedge\Gamma_1] \\ \Pi_1 \\ A\end{array} \quad \ldots \quad \begin{array}{c}[\wedge\Gamma_n] \\ \Pi_n \\ A\end{array}}{(\wedge\Gamma_1) \vee \ldots \vee (\wedge\Gamma_n)}}{A} \vee\mathrm{E}$$

We can therefore conclude that $A \in \mathrm{TH}_{\mathrm{OM}}(O)$. ∎

TABLE 1. Metamodels and object theorems generated by a set of reflection rules

(RR)	Models of $\mathrm{TH}_{\mathrm{OM}}(M)$	$\mathrm{TH}_{\mathrm{OM}}(O)$
$\emptyset$	Any set of L_O-wffs	$\mathrm{TH}(O)$
$\mathrm{Rdw}_{\mathrm{r}}$	Any set of L_O-wffs	$\mathrm{TH}(O + \bigcap \mathfrak{M})$
$\mathrm{Rup}_{\mathrm{r}}$	$\mathrm{TH}(O) \subseteq m$	$\mathrm{TH}(O)$
Rdw	Any set of L_O-wffs	$\mathrm{TH}(O + \bigcap \mathfrak{M})$
Rup	$\mathrm{TH}(O) \subseteq m$	$\mathrm{TH}(O)$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$	Fixpoint equation (3.2)	$\bigcap \mathfrak{M}$
$\mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}} + \mathrm{Rdw}$	m is consistent with $\mathrm{TH}(O)$	$\mathrm{TH}(O + \bigcap \mathfrak{M})$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$	Any *extension* of O	$\bigcap \mathfrak{M}$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}}$	Any *consistent extension* of O	$\bigcap \mathfrak{M}$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}_{\mathrm{r}}^{\mathrm{n}}$	fixpoint equation (3.10)	$\bigcap \mathfrak{M}$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$	Any *consistent extension* of O	$\bigcap \mathfrak{M}$
$\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$	Any *maximal extension* of O	$\bigcap \mathfrak{M}$
$\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$	m is consistent with $\mathrm{TH}(O)$	$\bigcap_{m \in \mathfrak{M}} \mathrm{TH}(O + m)$
$\mathrm{Rup} + \mathrm{Rdw}$	Any *maximal extension* of O	$\bigcap \mathfrak{M}$
$\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}}$	Any *maximal consistent extension* of O	$\bigcap \mathfrak{M}$

Notice that, $\mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}} + \mathrm{Rdw}$ and $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$ generate the same metatheory but they do not generate the same object theory. Consider the following example.

EXAMPLE 3.14
Let OM be the OM pair composed of the empty object theory O and the metatheory M with axiom $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}q\text{”})$, connected by the reflection rules $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$. Let OM_r be the OM pair obtained from OM by adding the restriction to $\mathrm{Rup}^{\mathrm{n}}$. Since they have the same set of metatheorems, OM_r and OM have the same set of models $\mathfrak{M}$ defined as follows:

$$\mathfrak{M} = \{m \ : m \text{ is consistent and } p \in m \text{ or } q \in m\}. \tag{3.13}$$

The set of object theorems of OM_r differs from that of OM. Indeed, according to Theorem 3.13, we have the following. The set of object level theorems of OM_r is equal to the theorems provable from O extended with the intersection of all the elements of $\mathfrak{M}$, which is the empty set. Therefore, $\mathrm{TH}_{OM_r}(O) = \mathrm{TH}(O)$ is the set of the propositional tautologies. On the other hand the set of object level theorems of OM is equal to the intersection of the theorems provable from O extended with m for some $m \in \mathfrak{M}$. This means, for instance, that, since either p or q belongs to any m, $p \vee q$ is provable from $O + m$ for all m. This implies that $p \vee q \in \mathrm{TH}_{\mathrm{OM}}(O)$.

Table 1 summarizes the results proved in this section.

4 A strict ordering characterization

In Part I [2] reflection rules have been partially ordered; however, we still do not have explicit results stating that if $(\mathrm{RR})_i \leq_X (\mathrm{RR})_2$, it is not the case that $(\mathrm{RR})_i =_X (\mathrm{RR})_2$. In this section we show that all the combinations of reflection rules described in this paper generate

different object-meta relations, and therefore the partial order described in the previous paper can be transformed in a *strict partial order*.

Let us recall the definition of partial orders between combinations of reflection rules given in the previous paper.

DEFINITION 4.1
For each two OM pairs OM_1 and OM_2,

1. $\text{OM}_1 \leq_O \text{OM}_2$ if $\text{TH}_{\text{OM}_1}(O) \subseteq \text{TH}_{\text{OM}_2}(O)$;
2. $\text{OM}_1 \leq_M \text{OM}_2$ if $\text{TH}_{\text{OM}_1}(M) \subseteq \text{TH}_{\text{OM}_2}(M)$;
3. $\text{OM}_1 \leq_D \text{OM}_2$ if $\vdash_{\text{OM}_1} \subseteq \vdash_{\text{OM}_2}$;
4. $\text{OM}_1 =_X \text{OM}_2$ if $\text{OM}_1 \leq_X \text{OM}_2$ and $\text{OM}_2 \leq_X \text{OM}_1$, where $X \in \{O, M, D\}$;
5. $\text{OM}_1 <_X \text{OM}_2$ if $\text{OM}_1 \leq_X \text{OM}_2$ and not $\text{OM}_2 =_X \text{OM}_1$, where $X \in \{O, M, D\}$.

Orders on sets of reflection rules can be defined on the basis on the orders of OM pairs.

DEFINITION 4.2
Let $\text{OM}(\text{RR})$ be an OM pair composed of an object theory and a metatheory connected by the set of reflection rules (RR). For any combination of reflection rules $(\text{RR})_1$, $(\text{RR})_2$:

1. $(\text{RR})_1 \leq_O (\text{RR})_2$ if, for any $\text{OM}(\text{RR})_1$ and $\text{OM}(\text{RR})_2$, with the same object theory and metatheory, $\text{OM}(\text{RR})_1 \leq_O \text{OM}(\text{RR})_2$;
2. $(\text{RR})_1 \leq_M (\text{RR})_2$ if, for any $\text{OM}(\text{RR})_1$ and $\text{OM}(\text{RR})_2$ with the same object theory and metatheory, $\text{OM}(\text{RR})_1 \leq_M \text{OM}(\text{RR})_2$;
3. $(\text{RR})_1 \leq_D (\text{RR})_2$ if, for any $\text{OM}(\text{RR})_1$ and $\text{OM}(\text{RR})_2$ with the same object theory and metatheory, $\text{OM}(\text{RR})_1 \leq_D \text{OM}(\text{RR})_2$;
4. For any $X \in \{O, M, D\}$, $(\text{RR})_1 =_X (\text{RR})_2$, if both $(\text{RR})_1 \leq_X (\text{RR})_2$ and $(\text{RR})_2 \leq_X (\text{RR})_1$.
5. For any $X \in \{O, M, D\}$, $(\text{RR})_1 <_X (\text{RR})_2$, if $(\text{RR})_1 \leq_X (\text{RR})_2$ and not $(\text{RR})_2 =_X (\text{RR})_1$.

The orders defined above are not completely independent of one another. For instance, if two combinations of reflection rules are equivalent (i.e. $(\text{RR})_1 =_D (\text{RR})_2$) then they generate the same object theory and metatheory (i.e. $(\text{RR})_1 =_O (\text{RR})_2$ and $(\text{RR})_1 =_M (\text{RR})_2$). The relations among the three orders are summarized in Figure 1. In this figure, an arrow connects X to Y, if and only if $(\text{RR})_1 X (\text{RR})_2$ implies $(\text{RR})_1 Y (\text{RR})_2$. The correctness of this latter graph is straightforward.

THEOREM 4.3 (Correctness of the graph in Figure 2)
If the graph of Figure 2 contains an arc labelled with X from $(\text{RR})_1$ to $(\text{RR})_2$ then $(\text{RR})_1 X (\text{RR})_2$.

To prove the previous theorem we need the following lemma:

LEMMA 4.4 (Effective assumptions)
Let OM be an OM pair with the set (RR) of reflection rules. Let Γ_O, A_O and Γ_M, A_M be a set of wffs of L_O and L_M respectively. Then:

1. if the reflection up rules in (RR) are restricted, then $\Gamma_O, \Gamma_M \vdash_{\text{OM}} A_M$ if and only if $\Gamma_M \vdash_{\text{OM}} A_M$;

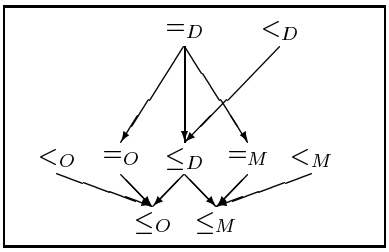

FIGURE 1. Relations among orders

2. if the reflection down rules in (RR) are restricted, then $\Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_O$ if and only if $\Gamma_O \vdash_{\mathrm{OM}} A_O$.

To prove (1), it is enough to observe that only way to 'switch' from object to meta in order to infer a metaformula from object assumptions, is by applying restricted reflection rules. However, these rules are applicable only if all the assumptions in the object theory are discharged. A similar argument can be used to prove (2).

PROOF. [Theorem 4.3] The graph in Figure 2 is obtained from the analogous graph in [2] by refining the relations between reflection rules. Since $<_O$ (resp. $<_M$ and $<_D$) are defined as the conjunction of $\leq_D$ and $\neq_O$ (resp. the conjunction of $\leq_D$ and $\neq_M$ and the conjunction of $\leq_D$ and $\neq_D$), the additional information encoded in the graph in Figure 2, with respect to the previous version, consists of statements asserting that two combinations of reflection rules differ either in the object and/or metatheory they generate, or in their derivability relation.

Following Definition 4.2, we prove $(\mathrm{RR})_1 \neq_O (\mathrm{RR})_2$ (resp. $(\mathrm{RR})_1 \neq_M (\mathrm{RR})_2$), by showing that there are two OM pairs OM_1 and OM_2 composed of the same object and metatheories and connected by $(\mathrm{RR})_1$ and $(\mathrm{RR})_2$ respectively, which generate a different object (resp. meta) theory. In practice, we look for an object (resp. meta) wff A not provable in OM_1 and provable in OM_2. $(\mathrm{RR})_1 \neq_D (\mathrm{RR})_2$ is automatically proved as a consequence of $(\mathrm{RR})_1 \neq_O (\mathrm{RR})_2$ or $(\mathrm{RR})_1 \neq_M (\mathrm{RR})_2$. We prove $\neq_D$ without proving $(\mathrm{RR})_1 \neq_O (\mathrm{RR})_2$ or $(\mathrm{RR})_1 \neq_M (\mathrm{RR})_2$ by providing two OM pairs OM_1 and OM_2 and a set of object and meta wffs Γ, A such that A is not derivable from Γ in OM_1, and A is derivable from Γ in OM_2.

We proceed by considering the arcs of the graph in Figure 2 from the bottom up. We say that a (meta or object) theory is an empty (meta or object) theory meaning that it has no axioms.

1. $\emptyset <_M \mathrm{Rup}_\mathrm{r}$. (We prove $\emptyset \neq_M \mathrm{Rup}_\mathrm{r}$.) Consider the OM pairs OM_1 and OM_2 composed of the empty object theory and the empty metatheory connected by no reflection rules and by Rup_r respectively. No wff of the form $\bullet(\text{“}A\text{”})$ is a theorem of the metatheory of OM_1, while, for instance, $\bullet(\text{“}\neg\bot\text{”})$ is a metatheorem of OM_2.
2. $\emptyset <_O \mathrm{Rdw}_\mathrm{r}$. (We prove $\emptyset \neq_O \mathrm{Rdw}_\mathrm{r}$.) Consider the OM pairs OM_1 and OM_2 composed of the empty object theory and the metatheory with the only axiom $\bullet(\text{“}\bot\text{”})$ connected by no reflection rules and by Rdw_r respectively. Since the object theory of OM_1 is equal to O, $\bot$ is not provable, Instead $\bot$ is provable (by Rdw_r) in the object theory of OM_2.
3. $\mathrm{Rup}_\mathrm{r} <_D \mathrm{Rup}$. (We prove $\mathrm{Rup}_\mathrm{r} \neq_D \mathrm{Rup}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty metatheory M, connected by Rup_r and Rup respectively. The metatheory of OM_1 is equal to M and therefore $\bullet(\text{“}\bot\text{”})$ is not

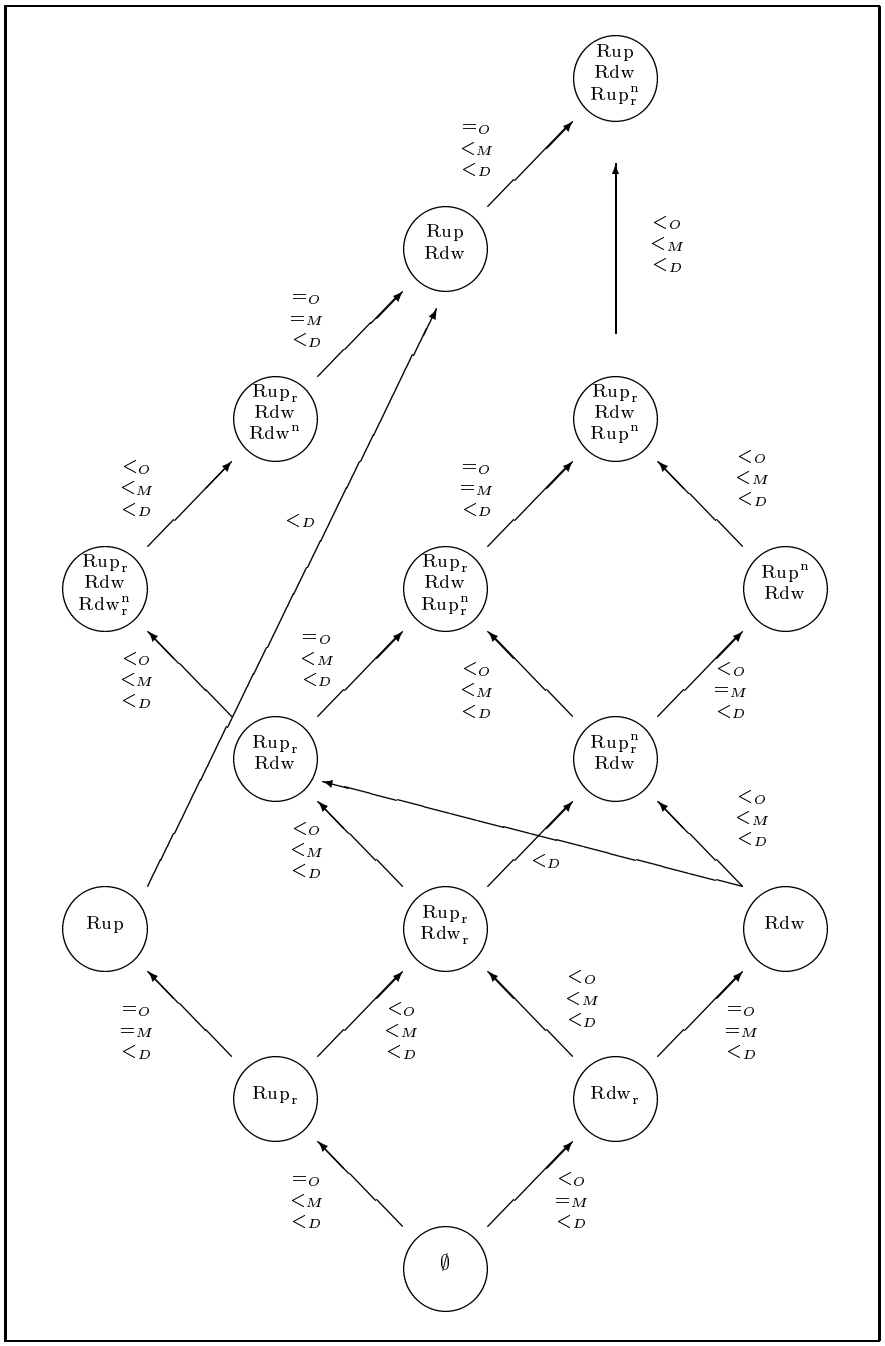

FIGURE 2. Strict ordering of (RR)s

provable in it. Furthermore, by Lemma 4.4, $\bullet(\text{“}\bot\text{”})$ is not derivable in the metatheory from the assumption $\bot$ of the object theory. In OM_2 instead, $\bullet(\text{“}\bot\text{”})$ is derivable (by Rup) from the assumption $\bot$ in the object theory.

4. $\text{Rdw}_\text{r} <_D \text{Rdw}$. Analogous to the previous case.
5. $\text{Rup}_\text{r} <_M \text{Rup}_\text{r} + \text{Rdw}_\text{r}$. (We prove $\text{Rup}_\text{r} \neq_M \text{Rup}_\text{r} + \text{Rdw}_\text{r}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the metatheory M with the axiom $\bullet(\text{“}\bot\text{”})$ connected by Rup_r and $\text{Rup}_\text{r} + \text{Rdw}_\text{r}$ respectively. By Theorem 3.1, the interpretation $m = \text{TH}(O) \cup \{\bot\}$ is a model of the metatheory of OM_1. As O is consistent, there is an object wff A distinct from $\bot$ such that $m \not\models \bullet(\text{“}A\text{”})$ and therefore $\bullet(\text{“}A\text{”})$ is not a theorem of the metatheory of OM_1. In OM_2, instead, each formula $\bullet(\text{“}A\text{”})$ is provable as follows:
$$\dfrac{\dfrac{\bullet(\text{“}\bot\text{”})}{\bot}\ \text{Rdw}_\text{r} \quad \dfrac{}{A}\ \bot}{\bullet(\text{“}A\text{”})}\ \text{Rup}_\text{r}$$
6. $\text{Rup}_\text{r} <_O \text{Rup}_\text{r} + \text{Rdw}_\text{r}$. (We prove $\text{Rup}_\text{r} \neq_O \text{Rup}_\text{r} + \text{Rdw}_\text{r}$.) In the above proof we can see that $\bot$ is provable in the object theory of OM_2 while it is not provable in the object theory of OM_1, because there is no rule going from the metatheory to the object theory.
7. $\text{Rdw}_\text{r} <_O \text{Rup}_\text{r} + \text{Rdw}_\text{r}$. (We prove $\text{Rdw}_\text{r} \neq_O \text{Rup}_\text{r} + \text{Rdw}_\text{r}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the metatheory M with the axiom $\neg\bullet(\text{“}\neg\bot\text{”})$ connected by Rdw_r and $\text{Rup}_\text{r} + \text{Rdw}_\text{r}$ respectively. From [2, Theorem 4.1] we have that the set of object theorems of OM_1 is $\text{TH}(O + \text{Rdw}(\text{TH}(M)))$. Since M does not prove any theorem of the form $\bullet(\text{“}A\text{”})$, the set of object theorems of OM_1 is $\text{TH}(O)$. In particular $\bot$ is not provable in O and therefore it is not provable in OM_1. In OM_2 instead $\bot$ is a theorem of the object theory. Indeed:
$$\dfrac{\dfrac{\dfrac{\dfrac{\bot}{\neg\bot}\ \supset\text{I}}{\bullet(\text{“}\neg\bot\text{”})}\ \text{Rup}_\text{r} \qquad \neg\bullet(\text{“}\neg\bot\text{”})}{\bot}\ \supset\text{E}}{\dfrac{\dfrac{\bot}{\bullet(\text{“}\bot\text{”})}\ \bot}{\bot}\ \text{Rdw}_\text{r}}$$
8. $\text{Rdw}_\text{r} <_M \text{Rup}_\text{r} + \text{Rdw}_\text{r}$. (We prove $\text{Rdw}_\text{r} \neq_M \text{Rup}_\text{r} + \text{Rdw}_\text{r}$.) Consider the OM pairs OM_1, and OM_2 composed by the empty object theory and metatheory connected by Rup_r and $\text{Rup}_\text{r} + \text{Rdw}_\text{r}$ respectively. $\bullet(\text{“}\top\text{”})$ is not provable in OM_1 as the empty set is a model of the metatheory of OM_1, while $\bullet(\text{“}\top\text{”})$ is provable in OM_2.
9. $\text{Rup} <_D \text{Rup} + \text{Rdw}$. (We prove $\text{Rup} \neq_D \text{Rup} + \text{Rdw}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty metatheory M, connected by Rup and $\text{Rup} + \text{Rdw}$ respectively. By [2, Theorem 4.1] we have that the set of theorems of the object theory of OM_1 is $\text{TH}(O)$ and therefore $\bot$ is not provable in the object theory of OM_1. Therefore, by Lemma 4.4, in OM_1 $\bot$ is not derivable in the object theory from $\bullet(\text{“}\bot\text{”})$. In OM_2, instead, $\bot$ can be derived in the object theory from $\bullet(\text{“}\bot\text{”})$ in the metatheory (via a single application of Rdw).
10. $\text{Rup}_\text{r} + \text{Rdw}_\text{r} <_M \text{Rup}_\text{r} + \text{Rdw}$. (We prove $\text{Rup}_\text{r} + \text{Rdw}_\text{r} \neq_M \text{Rup}_\text{r} + \text{Rdw}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty

metatheory M, connected by $\mathrm{Rup_r} + \mathrm{Rdw_r}$ and $\mathrm{Rup_r} + \mathrm{Rdw}$ respectively. By Theorem 3.3 the set of models of the metatheory of OM_1 is $\{m : \mathrm{TH}(O) \subseteq m\}$. Consider the model of the metatheory of OM_1 $m = \mathrm{TH}(O) \cup \{\bot\}$. Clearly $m \not\models \bullet(\text{“}\bot\text{”}) \supset \bullet(\text{“}p\text{”})$, and therefore $\bullet(\text{“}\bot\text{”}) \supset \bullet(\text{“}p\text{”})$ is not a theorem of OM_1. However, $\bullet(\text{“}\bot\text{”}) \supset \bullet(\text{“}p\text{”})$ is a theorem of OM_2.

11. $\mathrm{Rup_r} + \mathrm{Rdw_r} <_O \mathrm{Rup_r} + \mathrm{Rdw}$. (We prove $\mathrm{Rup_r} + \mathrm{Rdw_r} \neq_O \mathrm{Rup_r} + \mathrm{Rdw}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the metatheory M with axiom $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}q\text{”})$, connected by $\mathrm{Rup_r} + \mathrm{Rdw_r}$ and $\mathrm{Rup_r} + \mathrm{Rdw}$, respectively. By Theorem 3.3 the set of models of the metatheory of OM_1 is:

$$\{m : \mathrm{TH}(O) \subseteq m \text{ and } p \in m \text{ or } q \in m\}. \tag{4.1}$$

By Theorem 3.11 the set of theorems of the object theory of OM_1 is the intersection of m for each m in (4.1), which is $\mathrm{TH}(O)$. The formula $p \vee q$ is not provable in $\mathrm{TH}(O)$ as O, being empty, proves only classical tautologies. This implies that $p \vee q$ is not a theorem of the object theory of OM_1. In OM_2, instead, $p \vee q$ is provable as follows:

$$\dfrac{\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}q\text{”}) \quad \dfrac{\dfrac{\dfrac{\bullet(\text{“}p\text{”})}{p}\,\mathrm{Rdw}}{p \vee q}\,\vee\mathrm{I}}{\bullet(\text{“}p \vee q\text{”})}\,\mathrm{Rup_r} \quad \dfrac{\dfrac{\dfrac{\bullet(\text{“}q\text{”})}{q}\,\mathrm{Rdw}}{p \vee q}\,\vee\mathrm{I}}{\bullet(\text{“}p \vee q\text{”})}\,\mathrm{Rup_r}}{\bullet(\text{“}p \vee q\text{”})}\,\vee\mathrm{E}}{p \vee q}\,\mathrm{Rdw}$$

12. $\mathrm{Rdw} <_M \mathrm{Rup_r^n} + \mathrm{Rdw}$. (We prove $\mathrm{Rdw} \neq_M \mathrm{Rup_r^n} + \mathrm{Rdw}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty metatheory M. By [2, Theorem 4.1], the set of theorems of the metatheory of OM_1 is $\mathrm{TH}(M)$ (i.e. the set of tautologies of L_M), and by [2, Theorem 4.1], the set of metatheorems of OM_2 is $\mathrm{TH}(M + (\mathrm{Cons}))$ which contains wffs which are not tautologies of L_M (e.g. $\bullet(\text{“}p\text{”}) \wedge \bullet(\text{“}q\text{”}) \supset \neg \bullet (\text{“}\neg p \vee \neg q\text{”})$).

13. $\mathrm{Rdw} <_O \mathrm{Rup_r^n} + \mathrm{Rdw}$. (We prove $\mathrm{Rdw} \neq_O \mathrm{Rup_r^n} + \mathrm{Rdw}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the metatheory M with the axiom $\bullet(\text{“}\bot\text{”}) \vee \bullet(\text{“}p\text{”})$. By Theorem 3.1 the set of models of the metatheory of OM_1 is the set of models of M, i.e. the set of models m such that $\bot \in m$ or $p \in m$. By Theorem 3.11 the set of object theorems of OM_1 is the set of theorems of O extended with the intersection of the models of M. Since this intersection is empty, the object theory of OM_1 is equal to O and, therefore, it does not prove p. In the object theory of OM_2, instead, p is provable as follows:

$$\dfrac{\bullet(\text{“}\bot\text{”}) \vee \bullet(\text{“}p\text{”}) \quad \dfrac{\bullet(\text{“}\bot\text{”}) \quad \dfrac{\dfrac{\dfrac{\bot}{\neg\bot}\,\supset\mathrm{I}}{\neg \bullet (\text{“}\bot\text{”})}\,\mathrm{Rup_r^n}}{\bullet(\text{“}p\text{”})}\,\supset\mathrm{E}}{\bullet(\text{“}p\text{”})}\,\bot \qquad \bullet(\text{“}p\text{”})}{\dfrac{\bullet(\text{“}p\text{”})}{p}\,\mathrm{Rdw}}\,\vee\mathrm{E}$$

14. $\mathrm{Rup_r} + \mathrm{Rdw} <_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw_r^n}$. (We prove $\mathrm{Rup_r} + \mathrm{Rdw} \neq_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw_r^n}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object

theory O and the metatheory M, with the axiom $\neg \bullet (\text{“}p\text{”})$. Since p is not provable form the empty object theory, the interpretation $m = \mathrm{TH}(O)$ is a model of the metatheory (according to Theorem 3.2). Since also $\neg p$ is not provable in the empty object theory (i.e. $\neg p \notin \mathrm{TH}(O)$), $m \not\models \bullet(\text{“}\neg p\text{”})$ and therefore $\bullet(\text{“}\neg p\text{”})$ is not provable in the metatheory of OM_1. In OM_2 instead $\bullet(\text{“}\neg p\text{”})$ is provable. A proof of $\bullet(\text{“}\neg p\text{”})$ is the following:

$$\dfrac{\dfrac{\neg \bullet (\text{“}p\text{”})}{\neg p}\ \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}}{\bullet(\text{“}\neg p\text{”})}\ \mathrm{Rup}_{\mathrm{r}}$$

15. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$. (We prove $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} \neq_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$). Consider the two OM pairs of the previous case. The object theory of OM_1 does not prove $\neg p$, Indeed, by Theorem 3.1, the object theory of OM_1 is the intersection of the models of the metatheory of OM_1. The model m considered in the previous case is such that $\neg p \notin m$. Therefore $\neg p$ is not provable in the object theory of OM_1. In OM_2 instead $\neg p$ is provable (by applying $\mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$ to $\neg \bullet (\text{“}p\text{”})$).
16. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} <_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. (We prove that $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} \neq_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty metatheory M. The metatheory of OM_1 does not prove $\neg \bullet (\text{“}\bot\text{”})$. Indeed the model m which interprets $\bullet$ in the set of wffs L_O, is a model of $\mathrm{TH}_{\mathrm{OM}_1}(M)$. The metatheory of OM_2 instead proves $\neg \bullet (\text{“}\bot\text{”})$. A proof is the following.

$$\dfrac{\dfrac{\dfrac{\bot}{\neg\bot}\ \supset\mathrm{I}}{\neg \bullet (\text{“}\bot\text{”})}\ \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}} \tag{4.2}$$

17. $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_O \mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$. See Example 3.14.
18. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}} <_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$. (We prove $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}} \neq_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O and the empty metatheory M, connected by the reflection rules $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$ and $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$ respectively. By Theorem 3.9, the set of models of the metatheory of OM_1 is equal to the set of interpretations which are the sets of theorems of an extension of O. The interpretation $m = \mathrm{TH}(O)$ is, therefore, a model of the metatheory of OM_1. Since neither p nor $\neg p$ is provable in O, then neither $m \models \bullet(\text{“}p\text{”})$ nor $m \models \bullet(\text{“}\neg p\text{”})$ and therefore $m \not\models \bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$. This implies that $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$ is not valid, and therefore not provable, in the metatheory of OM_1. In the metatheory of OM_2, instead, $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$ is provable as follows:

$$\dfrac{\bullet(\text{“}p\text{”}) \vee \neg \bullet (\text{“}p\text{”}) \qquad \dfrac{\bullet(\text{“}p\text{”})}{\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})}\ \vee\mathrm{I} \qquad \dfrac{\dfrac{\dfrac{\neg \bullet (\text{“}p\text{”})}{\neg p}\ \mathrm{Rdw}^{\mathrm{n}}}{\bullet(\text{“}\neg p\text{”})}\ \mathrm{Rup}_{\mathrm{r}}}{\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})}\ \vee\mathrm{I}}{\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})}\ \vee\mathrm{E}$$

19. $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$. (We prove $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}} \neq_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the metatheory M with the axiom $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”}) \supset \bullet(\text{“}q\text{”})$, connected by the reflection rules $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}_{\mathrm{r}}$ and $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$

respectively. By Theorem 3.9 it can be proved that the set of models of the metatheory of OM_1 is equal to the set of interpretations m which are the set of theorems of an extension of O, and $q \in m$ if $p \in m$ or $\neg p \in m$. Since neither p nor $\neg p$ is provable in O, then the interpretation $m = \mathrm{TH}(O)$ is a model of the metatheory of OM_1. Theorem 3.11 implies that the set of theorems of the object theory of OM_1 is equal to the intersection of each model of the metatheory. Since q does not belong to a model m, then p is not provable in the object theory of OM_1. In OM_2, instead, q is provable as $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$ is a theorem of the metatheory of OM_2 (see the previous example) and q can be derived by $\supset$E and Rdw.

20. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} <_D \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$. (We prove $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} \neq_D \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the empty metatheory M, connected by the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$ and $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$ respectively. By Theorem 3.2 the metatheory of OM_1 has at least a model (e.g. $m = \mathrm{TH}(O)$), and therefore it is consistent. By Lemma 4.4, the non-derivability of $\bot$ in the metatheory implies the non-derivability of $\bot$ in the metatheory starting from $\bot$ in the object theory. This derivation is instead possible in OM_2:

$$\dfrac{\dfrac{\dfrac{\bot}{\neg\neg\bot}\,\bot}{\neg\bullet(\text{“}\neg\bot\text{”})}\,\mathrm{Rup^n} \qquad \dfrac{\dfrac{\bot}{\neg\bot}\,\supset\mathrm{I}}{\bullet(\text{“}\neg\bot\text{”})}\,\mathrm{Rup_r}}{\bot}\,\supset\mathrm{E}$$

21. $\mathrm{Rup^n} + \mathrm{Rdw} <_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$. (We prove $\mathrm{Rup^n} + \mathrm{Rdw} \neq_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the empty metatheory M, connected by the reflection rules $\mathrm{Rup^n} + \mathrm{Rdw}$ and $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$ respectively. By Theorem 3.2, the interpretation $m = \emptyset$ is a model of the metatheory of OM_1. This model does not satisfy $\bullet(\text{“}\neg\bot\text{”})$. This implies that $\bullet(\text{“}\neg\bot\text{”})$ is not a theorem of the metatheory of OM_1. In OM_2 instead, $\bullet(\text{“}\neg\bot\text{”})$ is provable (by $\mathrm{Rup_r}$).

22. $\mathrm{Rup^n} + \mathrm{Rdw} <_O \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$. (We prove $\mathrm{Rup^n} + \mathrm{Rdw} \neq_O \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the metatheory M with axiom $\bullet(\text{“}\neg\bot\text{”}) \supset \bullet(\text{“}\bot\text{”})$, connected by the reflection rules $\mathrm{Rup^n} + \mathrm{Rdw}$ and $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$ respectively. By Theorem 3.13 we have that the set of theorems of the object theory of OM_1 is equal to the intersection of the sets $\mathrm{TH}(O + m)$ for each model m of M. According to Theorem 3.2 the intrepretation $m = \emptyset$, is a model of the metatheory of OM_1. This implies that the set of theorems of the object theory of OM_1 is $\mathrm{TH}(O)$. Note that O is empty and therefore consistent. The object theory of OM_2, instead, is inconsistent, as $\bot$ is provable in the object theory of OM_2.

23. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n} <_D \mathrm{Rup} + \mathrm{Rdw}$. (We prove $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n} \neq_D \mathrm{Rup} + \mathrm{Rdw}$.) The reasoning is similar to that done for the case $\mathrm{Rup_r} <_D \mathrm{Rup}$.

24. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n} <_M \mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup_r^n}$. (We prove that $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n} \neq_M \mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup_r^n}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the empty metatheory M, connected by the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$ and $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup_r^n}$ respectively. According to Theorem 3.2 the interpretation $m = \mathrm{TH}(O)$ is a model of the metatheory of OM_1. Note that, for any proposition p, neither p nor $\neg p$ is provable in O. From this we can infer that $m \not\models \bullet(\text{“}p\text{”})$

and $m \not\models \bullet(\text{“}\neg p\text{”})$, i.e. $m \not\models \bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$. This implies that $\bullet(\text{“}p\text{”}) \vee \bullet(\text{“}\neg p\text{”})$ is not provable in the metatheory of OM_1. The same formula, however, is provable in the metatheory of OM_2 (see proof above).

25. $\mathrm{Rup} + \mathrm{Rdw} <_M \mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. (We prove that $\mathrm{Rup} + \mathrm{Rdw} \neq_M \mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$.) Consider two OM pairs OM_1 and OM_2 composed of the empty object theory O, and the empty metatheory M, connected by the reflection rules $\mathrm{Rup} + \mathrm{Rdw}$ and $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$, respectively. According to Theorem 3.2 the interpretation $mm = L_O$ is a model of the metatheory of OM_1. Since this model does not satisfy $\neg \bullet (\text{“}\bot\text{”})$, $\neg \bullet (\text{“}\bot\text{”})$ is not provable in the metatheory of OM_1. As showed before $\neg \bullet (\text{“}\bot\text{”})$ is a theorem of the metatheory of OM_2.
26. $\mathrm{Rdw} <_D \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$. Note that $\mathrm{Rdw} =_O \mathrm{Rdw}_{\mathrm{r}}$ and $\mathrm{Rdw}_{\mathrm{r}} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$, implies that $\mathrm{Rdw} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}}$. From the fact that $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}_{\mathrm{r}} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$, we have that $\mathrm{Rup} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$. Since $(\mathrm{RR})_1 \leq_D (\mathrm{RR})_2$ and $(\mathrm{RR})_1 <_O (\mathrm{RR})_2$ implies $(\mathrm{RR})_1 <_D (\mathrm{RR})_2$, we conclude that $\mathrm{Rup} <_D \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw}$.
27. $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. Suppose, by contradiction, that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} =_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. From the fact that $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} =_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$, we have that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} =_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$. However, from the fact that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_O \mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$, and from the fact that $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$, we have that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$ and, therefore, that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} \neq_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$. This contradicts what we have proved before. We conclude therefore that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} \neq_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$.
28. $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. We reason by contradiction as before. Suppose that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} =_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. Then from the fact that $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} =_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$, we have that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} =_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$. On the other hand from the fact that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} =_M \mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$ and that $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw} <_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$, we have that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} <_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$ and therefore that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} \neq_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}$. This contradicts what we have proved before. We conclude therefore that $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw} \neq_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$.

∎

Further relations between sets of reflection rules can be inferred from the structure of the graph in Figure 2. Indeed we can exploit the following properties.

PROPOSITION 4.5
Let X and Y be two relational symbols describing a relation among two sets of reflection rules. Let x stand for D, M, or O. Let $(\mathrm{RR})_1$, $(\mathrm{RR})_2$ and $(\mathrm{RR})_3$ be sets of reflection rules. Then the following properties hold:

IMPLICATION If $(\mathrm{RR})_1 X (\mathrm{RR})_2$, and Y is reachable from X in graph of Figure 1, then $(\mathrm{RR})_1 Y (\mathrm{RR})_2$.

TRANSITIVITY-1 If $(\mathrm{RR})_1 \leq_x (\mathrm{RR})_2$ and $(\mathrm{RR})_2 \leq_x (\mathrm{RR})_3$, then $(\mathrm{RR})_1 \leq_x (\mathrm{RR})_3$.

TRANSITIVITY-2 If $(\mathrm{RR})_1 <_x (\mathrm{RR})_2$ and $(\mathrm{RR})_2 \leq_x (\mathrm{RR})_3$, then $(\mathrm{RR})_1 <_x (\mathrm{RR})_3$.

IDENTITY $(\mathrm{RR})_1 =_x (\mathrm{RR})$.

SUBSTITUTIVITY If $(\mathrm{RR})_1 =_x (\mathrm{RR})_2$ and $(\mathrm{RR})_2 Y_x (\mathrm{RR})_3$, then $(\mathrm{RR})_1 Y_X (\mathrm{RR})_3$.

These properties follow immediately from the definitions of $=_x, \leq_x$ and $<_x$, with $x \in \{O, M, D\}$. Table 2 links the holding of $=_D, =_O, =_M, <_D, <_O, <_M$ of any two sets of

TABLE 2. Strict ordering among (RR)s

	1 $\emptyset$	2 $\mathrm{Rdw_r}$	3 $\mathrm{Rup_r}$	4 Rdw	5 Rup	6 $\mathrm{Rup_r}$ $\mathrm{Rdw_r}$	7 $\mathrm{Rup_r^n}$ Rdw	8 $\mathrm{Rup_r}$ Rdw	9 $\mathrm{Rup_r}$ Rdw $\mathrm{Rup_r^n}$	10 $\mathrm{Rup_r}$ $\mathrm{Rdw_r^n}$	11 $\mathrm{Rup_r}$ Rdw $\mathrm{Rup^n}$	12 $\mathrm{Rup_r}$ Rdw $\mathrm{Rdw^n}$	13 $\mathrm{Rup^n}$ Rdw	14 Rup Rdw	15 Rup Rdw $\mathrm{Rup_r^n}$
1 $\emptyset$	$=_D$	$<_O$ $=_M$ $<_D$	$=_O$ $<_M$ $<_D$	$<_O$ $=_M$ $<_D$	$=_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
2 $\mathrm{Rdw_r}$	$=_M$	$=_D$	$<_M$	$=_O$ $=_M$ $<_D$	$<_M$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
3 $\mathrm{Rup_r}$	$=_O$	$<_O$	$=_D$	$<_O$	$=_O$ $=_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
4 Rdw	$=_M$	$=_O$ $=_M$	$<_M$	$=_D$	$<_M$	$<_O$ $<_M$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
5 Rup	$=_O$	$<_O$	$=_O$ $=_M$	$<_O$	$=_D$	$<_O$ $<_M$	$<_O$	$<_O$ $<_M$	$<_O$ $<_M$	$<_O$ $<_M$	$<_O$ $<_M$	$<_O$ $<_M$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
6 $\mathrm{Rup_r}$ $\mathrm{Rdw_r}$						$=_D$		$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$		$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
7 $\mathrm{Rup_r^n}$ Rdw							$=_D$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$	$<_O$ $=_M$ $<_D$	$<_O$	$<_O$ $<_M$ $<_D$
8 $\mathrm{Rup_r}$ Rdw								$=_D$	$=_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$	$=_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$		$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
9 $\mathrm{Rup_r}$ Rdw $\mathrm{Rup_r^n}$								$=_O$	$=_D$	$<_O$	$=_O$ $=_M$ $<_D$	$<_O$		$<_O$	$<_O$ $<_M$ $<_D$
10 $\mathrm{Rup_r}$ Rdw $\mathrm{Rdw_r^n}$										$=_D$		$<_O$ $<_M$ $<_D$		$<_O$ $<_M$ $<_D$	$<_O$ $<_M$ $<_D$
11 $\mathrm{Rup_r}$ Rdw $\mathrm{Rup^n}$								$=_O$	$=_O$ $=_M$		$=_D$	$<_O$		$<_O$	$<_O$ $<_M$ $<_D$
12 $\mathrm{Rup_r}$ Rdw $\mathrm{Rdw^n}$												$=_D$		$=_O$ $=_M$ $<_D$	$=_O$ $<_M$ $<_D$
13 $\mathrm{Rup^n}$ Rdw							$=_M$	$<_O$	$<_O$ $<_M$	$<_O$	$<_O$ $<_M$ $<_D$	$<_O$	$=_D$	$<_O$	$<_O$ $<_M$ $<_D$
14 Rup Rdw												$=_O$ $=_M$		$=_D$	$=_O$ $<_M$ $<_D$
15 Rup Rdw $\mathrm{Rup_r^n}$												$=_O$		$=_O$	$=_D$

reflection rules. A relational symbol X occurs in row $(\mathrm{RR})_1$ column $(\mathrm{RR})_2$ if and only if $(\mathrm{RR})_1 X (\mathrm{RR})_2$. The results in Table 2 can be derived from the facts explicitly stated in the graph in Figure 2 plus the properties IMPLICATION, TRANSITIVITY-1, TRANSITIVITY-2, IDENTITY and SUBSTITUTIVITY.

EXAMPLE 4.6
In order to show that $\mathrm{Rdw} <_M \mathrm{Rup_r} + \mathrm{Rdw_r}$, which is not implicitly stated in the graph of Figure 2 by a labelled arc, we reason as follows:

1. $\mathrm{Rdw_r} <_M \mathrm{Rup_r} + \mathrm{Rdw_r}$ is stated in the graph of Figure 2;
2. $\mathrm{Rdw} =_M \mathrm{Rdw_r}$ is stated in the graph of Figure 2;
3. $\mathrm{Rdw} <_M \mathrm{Rup_r} + \mathrm{Rdw_r}$ is derivable from 1. and 2. by SUBSTITUTIVITY.

THEOREM 4.7 (Completeness of the graph in Figure 2)
The graph in Figure 2 is *complete*. That is, for each two sets of reflection rules $(\mathrm{RR})_1$ and $(\mathrm{RR})_2$ and for each relational symbol $X \in \{=_D, =_O, =_M, <_D, <_O, <_M\}$, $(\mathrm{RR})_1 X (\mathrm{RR})_2$ iff $(\mathrm{RR})_1 X (\mathrm{RR})_2$ is derivable from the facts explicitly stated in the graph in Figure 2, and from IMPLICATION, TRANSITIVITY-1, TRANSITIVITY-2, IDENTITY and SUBSTITUTIVITY.

The proof of Theorem 4.7 can be done by showing that, if $<_X$ $[=_X]$ does not occur in the

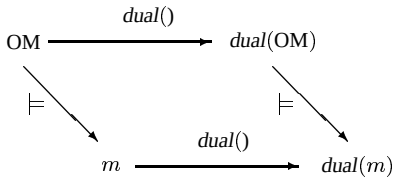

FIGURE 3. A duality graph for satisfiability

slot at row $(\mathrm{RR})_1$ column $(\mathrm{RR})_2$ of Table 2, then there are two OM pairs OM_1 and OM_2, such that $\mathrm{OM}_1 \not<_X \mathrm{OM}_2$ [$\mathrm{OM}_1 \neq_X \mathrm{OM}_2$]. The proof, not reported because it is routine and tedious, exploits the same techniques exploited in the proof of Theorem 4.3.

Let us discuss, top to bottom, left to right, some of the results in Table 2. The results in row 1 ($(\mathrm{RR})_1 = \emptyset$) are obvious. It is sufficient to notice that a single reflection up rule does not modify the object theory. Trivially, the dual result holds in the case of reflection down only. The results in rows 2 and 3 have similar explanations. The relations generated by a reflection up rule cannot be compared with those generated by a reflection down rule. A restricted rule generates the same meta and object theories as its corresponding unrestricted rule. It only allows more derivations, all those which (start in one theory and finish in another and) contain at least an application of a reflection rule in presence of open assumptions. The effects of dropping the restriction appear when we have at least one reflection up rule and one reflection down rule. In row 4 the only non-trivial result is in column 6. As from row 3 column 4, Rdw and Rdw_r generate the same meta and object theories. Adding Rup_r to Rdw_r increases the theorems of the metatheory, this proves $\mathrm{Rdw} <_M \mathrm{Rup}_\mathrm{r} + \mathrm{Rdw}_\mathrm{r}$. To prove that $\mathrm{Rdw} <_O \mathrm{Rup}_\mathrm{r} + \mathrm{Rdw}_\mathrm{r}$ it is sufficient to consider the case where $\Omega_O = \emptyset$ and $\Omega_M = \bullet(\text{“}A \vee \neg A\text{”}) \supset \bullet(\text{“}B\text{”})$. Finally, to prove $\mathrm{Rdw} \not<_D \mathrm{Rup}_\mathrm{r} + \mathrm{Rdw}_\mathrm{r}$ it is sufficient to further consider the fact that any derivation which contains an application of reflection down with open assumptions is not a derivation of the OM pair with $\mathrm{Rup}_\mathrm{r} + \mathrm{Rdw}_\mathrm{r}$. In row 5 the result in column 6 can be proved with arguments similar to those for row 4, column 6.

5 A duality property for satisfiability

As already stated in Part I [2], duality principles usually state the preservation of certain logical properties, e.g. provability or satisfiability, under appropriate syntactic transformations on formulas. Examples of duality principles for propositional logic and modal logic can be found in [5] and [1] respectively. In this section we consider satisfiability, in particular we prove (under certain general conditions) the correctness of the graph in Figure 3, where *dual*(.) is a transformation (defined below) on formulas/ models and OM pairs.

We recall from Part I [2] the notions of dual reflection rule, dual formula, and dual OM pair. See Part I [2] for an explanation of the underlying intuitions.

DEFINITION 5.1 (Dual reflection rule)
For any positive [negative] reflection rule ρ, *dual*(ρ) is the corresponding negative [positive]

reflection rule.

DEFINITION 5.2 (Dual formula)
For any formula $A \in L_M$, *dual*(A) is defined accordingly to 1–4:

1. *dual*$(A) = A$, A does not contain any occurrence of $\bullet$;
2. *dual*$(\bullet(\text{“}A\text{”})) = \neg \bullet (\text{“}\neg A\text{”})$;
3. *dual*$(A \circ B) =$ *dual*$(A) \circ$ *dual*(B), where $\circ$ is either $\wedge$ or $\vee$ or $\supset$;
4. *dual*$(\neg A) = \neg$*dual*(A).

Notice that, *dual*$(.)$ only modifies formulas containing $\bullet$ predicate, and preserves the structure of the connectives. This is not a surprise, as we are interested in a duality principle for metaproperties. To define *dual*$(\bullet(\text{“}A\text{”}))$ we reason as follows: since $\bullet(\text{“}A\text{”})$ is derivable by Rup from A, *dual*$(\bullet(\text{“}A\text{”}))$ must be derivable by *dual*(Rup) (i.e. $\mathrm{Rup}^{\mathrm{n}}$) from A. This implies that *dual*$(\bullet(\text{“}A\text{”}))$ must be an element of the set of formulas derivable from A by $\mathrm{Rup}^{\mathrm{n}}$. In symbols *dual*$(\bullet(\text{“}A\text{”})) \in \{\neg \bullet (\text{“}B\text{”}) : A \vdash_{\mathrm{o}} \neg B\}$. We have chosen *dual*$(\bullet(\text{“}A\text{”})) = \neg \bullet (\text{“}\neg A\text{”})$.

An important property of dualization is that the dual of the dual of an entity (formula, OM pair) is the entity itself (or equivalent, giving some appropriate notion of equivalence). To satisfy this requirement, if A is of the form $\bullet(\text{“}B\text{”})$, $\bullet(\text{“}B\text{”})$ must be equivalent to $\neg\neg \bullet (\text{“}\neg\neg B\text{”})$, which is *dual*$($*dual*$(A))$. This is not true in all OM pairs. We therefore restrict ourselves only to those OM pairs which satisfy the following property:

$$\vdash_{\mathrm{OM}} \bullet(\text{“}A\text{”}) \equiv \bullet(\text{“}\neg\neg A\text{”}). \tag{5.1}$$

We call the OM pairs which satisfy Condition (5.1), *classical*.

DEFINITION 5.3 (Dual OM pair)
If $\mathrm{OM} = \langle O, M, (\mathrm{RR})\rangle$ is classical OM pair, then the *dual* of OM, is the OM pair *dual*$(\mathrm{OM}) = \langle O,$ *dual*$(M),$ *dual*$((\mathrm{RR}))\rangle$, where, *dual*(M), is the theory obtained by substituting the axioms Ω_M of M with *dual*(Ω_M) and by adding $\bullet(\text{“}A\text{”}) \equiv \bullet(\text{“}\neg\neg A\text{”})$.

Let us define the dualization for the interpretations of the language of the metatheory. We require that an interpretation m satisfies $\bullet(\text{“}A\text{”})$ if and only if the dual interpretation of m satisfies $\neg \bullet (\text{“}\neg A\text{”})$. This implies that:

$$A \in m \iff \neg A \notin \textit{dual}(m). \tag{5.2}$$

By replacing m with *dual*(m) and A with $\neg A$ in (5.2) we obtain:

$$\neg A \notin \textit{dual}(m) \iff \neg\neg A \in \textit{dual}(\textit{dual}(m)). \tag{5.3}$$

Combining (5.2) and (5.3) we have that

$$A \in m \iff \neg\neg A \in \textit{dual}(\textit{dual}(m)). \tag{5.4}$$

Finally we require that *dual*$($*dual*$(m))$ satisfies the same formulas as m (this is in order to satisfy the principle that the dual of the dual of an entity is the entity itself). Hence from (5.4) we can conclude:

$$A \in m \iff \neg\neg A \in m. \tag{5.5}$$

Notice that restriction (5.5) is the semantical counterpart of restriction (5.1) for classical OM pairs. In the following, *dual*$(.)$ is defined only for those interpretations which satisfy condition (5.5). We are now ready to give the notion of dual interpretation.

DEFINITION 5.4 (Dual interpretation)
For each interpretation m of the metalanguage that satisfies condition (5.5), the interpretation *dual*$(m) = \{A : \neg A \notin m\}$.

A dual interpretation preserves satisfiability.

THEOREM 5.5 (Duality preserves satisfiability)
For each interpretation m of the metalanguage that satisfies condition (5.5), $m \models A$, if and only if *dual*$(m) \models$ *dual*(A).

PROOF. We proceed by induction on the complexity of the meta wffs. Let us start by proving that $m \models \bullet(\text{“}A\text{”})$, if and only if *dual*$(m) \models \neg \bullet (\text{“}\neg A\text{”})$. If $m \models \bullet(\text{“}A\text{”})$, then $A \in m$. By condition (5.5) $\neg\neg A \in m$. By definition of dual interpretation $\neg A \notin$ *dual*(m) and therefore *dual*$(m) \models \neg \bullet (\text{“}\neg A\text{”})$. Vice versa, if *dual*$(m) \models \neg \bullet (\text{“}\neg A\text{”})$, then $\neg A \notin$ *dual*(m), then, by definition of dual interpretation, $\neg\neg A \in m$. Condition (5.5) implies that $A \in m$ and therefore $m \models \bullet(\text{“}A\text{”})$. The step cases are routine as *dual*$(.)$ distributes over the connectives. ∎

Part I [2] shows that, for classical OM pairs, duality also preserves derivability, that is that:

$$\Gamma_O, \Gamma_M \vdash_{\text{OM}} A_O \iff \Gamma_O, \textit{dual}(\Gamma_M) \vdash_{\textit{dual}(\text{OM})} A_O \tag{5.6}$$

$$\Gamma_O, \Gamma_M \vdash_{\text{OM}} A_M \iff \Gamma_O, \textit{dual}(\Gamma_M) \vdash_{\textit{dual}(\text{OM})} \textit{dual}(A_M) \tag{5.7}$$

where Γ_O, A_O and Γ_M, A_M are a set of object wffs and a set of meta wffs respectively. Combining this latter result with Theorem 5.5 we obtain that the class of models of the metatheory generated by the OM pair *dual*(OM) is the dual of the class of models of OM. More formally:

THEOREM 5.6
For any classical OM pair OM, an interpretation m is a model of the metatheory of OM if and only if *dual*(m) is a model of the metatheory of *dual*(OM).

PROOF. Let m be a model of OM and A a theorem of the metatheory of *dual*(OM). By [2, Theorem 5.5], *dual*(A) is a theorem of OM and therefore $m \models$ *dual*(A). By (5.7), *dual*$(m) \models$ *dual*(*dual*(A)). The fact that *dual*(*dual*(A)) is equivalent to A ensures that *dual*$(m) \models A$. ∎

Theorem 5.6 allows us to extend the results about the models of the metatheory proved in the previous section. The generalization can be done for the set of reflection rules which are the dual of the combinations that generate classical OM pairs. We summarize these results in the following table, constructed from Table 1.

(RR)	Models of $\text{TH}_{\text{OM}}(M)$
$\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw}^{\text{n}}$	The set of formulas consistent with an extension of O
$\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw}^{\text{n}} + \text{Rup}_{\text{r}}$	The set of formulas consistent with a consistent extension of O
$\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw}^{\text{n}} + \text{Rdw}_{\text{r}}$	Fixpoint equation dual of (3.10)
$\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw}^{\text{n}} + \text{Rup}$	The set of formulas consistent with a consistent extension of O
$\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw}^{\text{n}} + \text{Rdw}$	Any consistent maximal extension of O or the empty set
$\text{Rup}^{\text{n}} + \text{Rdw}^{\text{n}}$	Any consistent maximal extension of O or the empty set
$\text{Rup}^{\text{n}} + \text{Rdw}^{\text{n}} + \text{Rup}_{\text{r}}$	Any consistent maximal extension of O

6 Conclusion

This paper follows from Part I [2], which investigates the proof theory of OM pairs. The core result of this paper is the definition of a model theory where the models of the metatheory, defined as sets of formulas, are put in correspondence with the object theory, and vice versa. This paper extends the results presented in Part I [2]. In particular, the partial ordering of OM pairs defined in Part I [2] is strengthened to a strict ordering, and the principle of duality for derivability introduced in Part I [2] is extended to satisfiability.

OM pairs are our proposed formalism for studying the foundations of metareasoning. OM pairs separately represent the three main factors of metalevel reasoning (i.e. the object theory, the metatheory, and the reflection principles) using three distinct components, namely, O, M and (RR). In many approaches a fourth factor is considered, namely, the amalgamation of the object and metatheory. The object theory and the metatheory can be considered as a single theory, or the metatheory can be embedded in the object theory, or the metatheory can contain the object theory. We have shown in Part I [2] that OM pairs can also formalize this aspect by using bridge rules of the form $\frac{M : A}{O : A}\ M \subseteq O$, and $\frac{O : A}{M : A}\ O \subseteq M$.

The first advantage of using OM pairs is that it is possible to reconstruct and classify many different and heterogeneous forms of metareasoning in terms of these four parameters. Examples on how different approaches can be reconstructed and compared using OM pairs, are given in the related work section of [2]. To support this comparison, we have studied the proof and model theoretic properties of many possible parameter configurations. For most of the combinations of reflection rules and kinds of object and metatheories (empty, horn, etc.), we have provided an axiomatic characterization, a semantic characterization and a Natural Deduction calculus.

A second advantage is that the explicit representation of these four components allows us to study the effects of each single component on the overall system, for instance, on the theorems or the models of the object theories and metatheories. To this extent, as an example, we have studied the effects of embedding the metatheory into the object theory, when they are connected by different reflection principles. We have proved that amalgamating them does not lead to inconsistent theories, ... and so on.

In order to complete our study of OM pairs, there is at least a further topic that must be accomplished. In these two papers, indeed, we limit ourself by considering reflection rules on *closed* formulas. There are, however, interesting reflection principles stating the correctness of the metatheory with respect to properties of the object theory on open formulas or terms. Examples of such properties are: *'being an open formula that represents a certain subset of the domain'*, *'being a term of the object theory that denotes a certain individual'*, ... Reflection principles on these properties have been studied in Formal Logics as *non-local reflection principles*, in Meta Logic Programming as *non-ground representations*, while in Knowledge Representation they are often tightly connected to the problem of *quantifying in*. To formalize these more general reflection principles, we have to consider reflection rules on open formulas. OM pairs with reflection rules defined on non-closed formulas will be the subject of further studies. However, it is our opinion that many of the results described in this paper will scale to these new kinds of reflection rules.

Acknowledgements

The research described in this paper owes a lot to the openness and sharing of ideas which exists in the Automated Reasoning Area at ITC-IRST and in the Mechanized Reasoning Group of the University of Trento and Genoa. Discussions with Massimo Benerecetti, Paolo Bouquet, Alessandro Cimatti, Chiara Ghidini, Enrico Giunchiglia, Silvio Imbó and Paolo Traverso, have provided useful insights. We also thank the anonymous reviewer for very useful feedback.

References

[1] B. F. Chellas. *Modal Logic – an Introduction*. Cambridge University Press, 1980.

[2] G. Criscuolo, F. Giunchiglia, and L. Serafini. A foundation for metareasoning, Part I: The proof theory. *Journal of Logic and Computation*, **12**, 167–208, 2002.

[3] S. Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, **27**, 259–316, 1962.

[4] F. Giunchiglia. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine*, XVI:345–364, 1993. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, pp. 39–49. Also IRST-Technical Report 9211-20, IRST, Trento, Italy, 1993.

[5] S.C. Kleene. *Introduction to Metamathematics*. North Holland, 1952.

[6] D. Prawitz. *Natural Deduction - A proof theoretical study*. Almquist and Wiksell, Stockholm, 1965.

Received 30 October 1997