A Foundation for Metareasoning
Part I: The Proof Theory

Giovanni Criscuolo[1] **Fausto Giunchiglia**[2,3]
Luciano Serafini[3]

[1]Dipartimento di Scienze Fisiche,
University of Naples, 80125 Napoli, Italy

[2]Dipartimento di Informatica e Studi Aziendali,
University of Trento, 38100 Trento, Italy

[3]ITC–IRST, Centro per la Ricerca Scientifica e Tecnologica,
38050 Trento, Italy

December 23, 2000

To be published in Journal of Logic and Computation.

Abstract

We propose a framework, called OM pairs, for the formalization of meta-reasoning. OM pairs allow us to generate deductively pairs composed of an object theory and a meta theory related via a so called reflection principle. This is done by imposing, via appropriate reflection rules, the relation we want to hold between the object theory and the meta theory. In this paper we concentrate on the proof theory of OM pairs. We study them from various points of view: we compare the strength of the object and the meta theories generated by different combination of reflection rules; for each combination we characterize the object theory and meta theory, both axiomatically (when possible), and by means of fix-point equations. Finally we study four important case studies.

1 Inroduction

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. To be about the object theory, the metatheory is defined starting from a language, called the *metalanguage*, that contains special formulas, called *metaformulas*, that represent properties of the object theory. Examples of properties which are often considered are: "being a variable (of the object theory)", "being a term", "being a formula", "a variable occurring free in a formula", "a formula being the result of substituting a term for all free occurrences of a variable in another formula", "being a theorem", "a formula being derivable from a set of formulas", "a formula being (in)consistent with respect to a set of formulas", "a formula being true in a model of the object theory", and so on.

Let us analyze in more detail the meaning of the statement "the meta theory is *about* the object theory". Intuitively a theory is a metatheory of an object theory if *the truth of metaformulas in the metatheory implies the holding of the corresponding properties of the object theory.* Formally, in order to say that a theory is meta of another, one should provide some formal framework to describe the object theory and the metatheory, and provide a statement, either in, or on such a framework, which corresponds to the intuitive statement above. In the literature, these kinds of statements are often called *reflection principles* [12]. A reflection principle for an object theory O is an assertion, either formal or informal, stating the correctness of the metatheory w.r.t. the object theory. In other words, a reflection principle is a (possibly formalized version of a) statement of the following form:

$$\textit{If the statement } \phi \textit{ is provable in a formal system } S \textit{ then } \phi \textit{ holds} \qquad (1)$$

Many other different reflection principles, similar to (1), can be found in the literature; see [18, 20] and [6] for a survey. The formalization of a reflection principle involves at least the three main aspects, namely: the language, the role, and the strength of the reflection principle itself. Let's analyze these three aspects in detail.

First, notice that (1) is an implication, where the premise is representable in the metalanguage of O, and the consequence is representable in O itself (as a matter of fact, it is ϕ itself). Statement (1) involves, however, a third (formal or informal) language, namely the language where (1) itself is stated. Depending on the choice of the languages where the premise, the conclusion and the statement itself is formulated we have different approaches. Examples where the object theory, the meta theory coincide and the reflection principle is stated in the same language are [12, 32, 6] (in formal logic) and [27, 28] (in Artificial Intelligence). In other approaches the object theory and the metatheory coincide but the reflection principle is stated externally to them, by means of a set of inference rules. In this case we have an informal representation of (1) which is not a formula in a formal system. Such examples are the linking rules defined in [8] (in Metalogic Programming) and the reflection rule of the FOL system [36]. Its clear that, if the object theory and the meta theory are completely distinct, then (1) cannot be stated in any of the two theories. Examples of such approaches are [36, 17] (in Artificial Intelligence) and [21, 22] (in Metalogic Programming).

A second observation about (1) concerns its role namely, how and to which purpose the reflection principle is used. A first possibility is that (1) has a *descriptive role.* This means that (1) is a statement that is true for a specific pair of object and metatheory. In this case the problem is to construct an object theory and a meta theory that meet the reflection principle. A second possibility is that (1) has a *generative role.* This means that the reflection principle, usually formalized as a set of axioms or a set of inference rules, is exploited to *generate* the object theory and the metatheory in the relation expressed by the reflection principle itself. Examples of the first approach can be found in the area of Formal Logics [12, 32, 6], Logic Programming [9], and Artificial Intelligence [27, 28]. Examples of reflection principle with a generative role can be found in the area of Artificial Intelligence [36, 17, 2, 26].

The third and last observation about (1) concerns its strength. A reflection principle can be stated for a single formula. This is the case, for instance, of the consistency statement where ϕ is the contradictory formula $\perp$. The same reflection principle can be stated for a restricted class of formulas, such as, for instance, the class of closed formulas (local reflection principle), the class of formulas of a certain complexity in the arithmetical hierarchy (partial reflections), the class of formulas which do not contain negation, and so on. In the area of Formal Logics, but also in Artificial Intelligence, one can find many results about the characterization of different reflection principles (see [6] for a survey).

Due to the heterogeneity on languages, roles and strength one can hardly compare, reuse, or compose the many different approaches. The goal of this paper is to provide a logical framework capable of representing all the different approaches by explicitly modeling the three aspects described above, namely: reflection principles in single and multiple theories; descriptive vs. generative role of the reflection principles; and finally, different forms of reflection principles.

Our work plan is as follows. We start by introducing a class of formal systems called *Object-Meta pairs* (OM pairs), which are composed of an object theory and a metatheory, connected by some reflection principle. To this extent we introduce a set of inference rules, called *Reflection Rules* (RR) that formally represent the most common reflection principles, and that can be used to derive facts by reasoning across the two theories (Section 2). For each meaningful combination of RRs, we study how the metatheory and the object theory are affected (Sections 3–5). We finally provide three case studies (Section 6), and compare OM pairs with the existing literature (Section 7). The core analysis of OM pairs is organized in three main steps. First, in Section 3, we compare the effects of different sets of reflection rules. We perform this analysis by comparing the strength of the derivability and provability relations generated. Second, in Section 4 we study the absolute strength of the OM pairs we consider. We show that certain combinations of reflection rules generate a metatheory that is also obtainable with the addition of well known metatheoretic axiom schemata. We also show that there are certain combinations of reflection rules for which such a set of axioms does not exist. Most interestingly, we show that this is the case of one of the most common and natural reflection principles. Finally, in Section 5 we define a notion of duality on reflection principles, which allows us to extend the results proved in the previous sections to new sets of reflection rules. Throughout the paper we consider OM pairs with propositional object theory and metatheory, but in Section 6.4 we generalize our analysis to OM pairs with first order object and metatheories.

2 OM pairs

2.1 The intuition

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). *Metaformulas*, namely formulas of L_M which refer to some property of O, are usually written using standard notation and exploiting two key features of the metalanguage. The first is the ability to refer to, or name, elements of the object theory, e.g., variables, terms, formulas, theorems, sets of formulas. One way to obtain this is by adding to the metalanguage a set of constants, often called *names*, such that for each name there is a corresponding element of the object theory. Notationally, we write names by surrounding the corresponding element with double quotes. Thus, for instance, "A" is the name of the sentential constant A. The second feature is the existence of a distinguished predicate, or *metapredicate*, written as "$\bullet$", which, intuitively, *represents*, in the metatheory, a corresponding object level property. Thus, for instance, if $\bullet$ is unary, then the intuition is that the holding of the metaformula $\bullet$("e") is dependent on the holding of e of the property being represented.

$\vdash_{\mathrm{O}}$ and $\vdash_{\mathrm{M}}$, 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 a new kind of inter-theory inference rules, called *bridge rules* [14], whose premises and conclusions belong to different languages. Thus we may have a bridge rule with premises in L_O and conclusion in L_M; and, vice versa, a bridge rule with premises in L_M and conclusion in L_O. Bridge rules extend deductively $\mathrm{TH}(O)$ and $\mathrm{TH}(M)$, that is, new object and meta theorems 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.

The intuitions underlying the use of bridge rules are the following:

1. a bridge rule (or a set of bridge rules) states a link between the holding of an object level property and the holding of a corresponding metaformula, namely, it states a *basic* Meta property.[1] Sets of bridge rules, state multiple Meta properties which, in some cases, may result in interesting *global* Meta properties;

2. a bridge rule, or a set of bridge rules, may be applied deductively and may generate otherwise unprovable meta and object theorems. The extra meta and object theorems are exactly those facts which must hold for the desired Meta property to hold. This idea is exemplified in Figure 1, where (RR) is a set of bridge rules. (RR) is essentially seen as an operator which maps a pair of theories O and M onto a new "appropriate" pair of theories $O' = \langle L_O, \Omega'_O, \Delta_O \rangle$ and $M' = \langle L_M, \Omega'_M, \Delta_M \rangle$, with $\mathrm{TH}(O) \subseteq \mathrm{TH}(O')$ and $\mathrm{TH}(M) \subseteq \mathrm{TH}(M')$.

Notice that this is somewhat opposite to what is the usual approach, where one directly defines meta and object theory (in some cases, e.g., PA, the object theory is

[1] This approach allows us to study only those properties of the object theory which can be defined in terms of the object theory derivability relation, e.g., being provable, being consistent, being a logical consequence. It does not allow us to study properties like: being a variable, being a term, being a well formed formulas (wff), a variable occurring free in a formula, and so on.

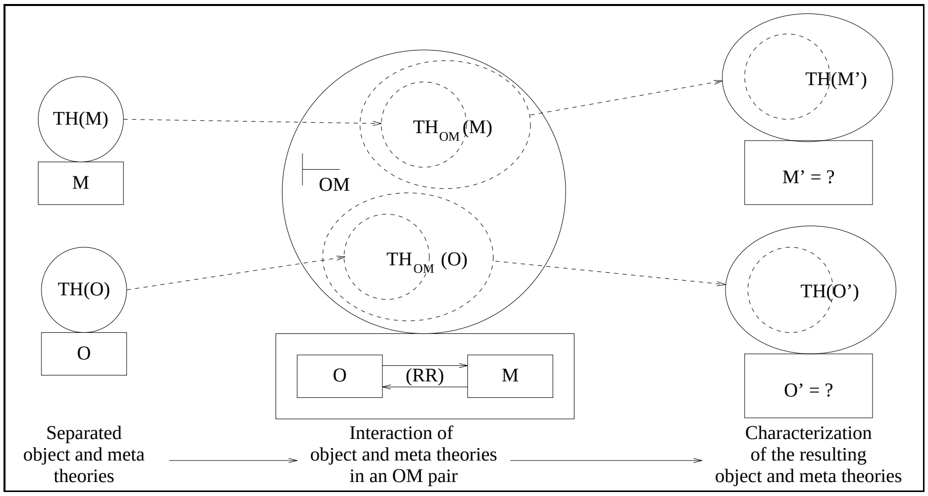

Figure 1: Meta and object theories deductively generated inside an OM pair.

expressive enough to be its own metatheory); and then verifies that the metatheory actually satisfies the desired Meta property. In the "standard" approach the process of generation of the metatheory is left implicit in the mind of the person defining it, while here the metatheory is deductively synthesized by imposing the desired Meta property via bridge rules. Our approach makes explicit the process of (object theory and) metatheory generation and, therefore, allows us to tune its parameters so that various, possibly very different, metatheories are generated.

The key step for our approach to be fruitful is that of defining "interesting" bridge rules. In principle any bridge rule is fine. However, in practice there are constraints that must be satisfied. First, each bridge rule should impose a basic Meta property whose meaning is intuitively clear, and possibly simple. Second, bridge rules should allow for the definition of metatheories with "interesting" global Meta properties. For instance they should allow us to define metatheories about provability, about derivability, about truth, and so on. Third, bridge rules should satisfy some (weak) criterion of minimality. That is, it should not be the case that a bridge rule can be trivially expressed as a combination of the others. We therefore focus on the following bridge rules:

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

For each of the bridge 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 $_r$), which can be applied under the following condition:

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

We call the above bridge rules, *reflection rules*, thus extending the terminology which can be found in the literature (see e.g., [36, 8, 18]). The Rup rules are called *reflection up* rules, the Rdw rules, *reflection down* rules. The rules with the r are called *restricted*, the others *unrestricted*; the rules with the n are called *negative*, the others *positive*.

Let us argue that our reflection rules meet the criteria described above. As shown in the rest of the paper, reflection rules allow us to define metatheories with interesting global Meta properties. Minimality and simplicity of the reflection rules is also true. In particular, the rules listed above are all natural variations of Rup and Rdw, the two reflection rules mainly studied in the literature (see again [36, 8, 18]). $\mathrm{Rup_r}$ and $\mathrm{Rdw_r}$ link provability in one theory with provability in the other (they can be applied only if there are no open assumptions). $\mathrm{Rup_r}$ establishes the completeness of the metatheory with respect the object theory, $\mathrm{Rdw_r}$, the correctness. All the reflection up rules establish some form of completeness, while all the reflection down rules establish some form of correctness. The unrestricted rules link derivability in one theory with provability in the other. The negative rules link the holding of a formula of form $\neg A$ with with the holding of $\neg \bullet (\text{“}A\text{”})$.

2.2 The formal notion

For the rest of the paper we make the following simplifying hypotheses: L_O and L_M are propositional and L_M is a *propositional metalanguage* of L_O. (In Section 6.4 we show how the result in the propositional case can be generalized to more complex cases. Propositional metalanguage is defined as follows.

Definition 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 (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 meta language of L_O, and (RR) is a set of reflection rules.

Given an object theory O, a meta theory 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 $\mathrm{Rup} + \mathrm{Rdw^n}$. We suppose that Δ_O and Δ_M contain the rules for classical propositional logic, as defined in [30].

An Object-Meta pair defines a derivability relation between multiple languages, which is a generalization of the usual single language consequence relation. A detailed definition of multilanguage derivability relation is given in [31]. In the following we give such a definition only for OM pairs. Roughly, the derivability relation defined by an OM pair is the transitive closure of the derivability relation of the object theory ($\vdash_{\mathrm{O}}$) and of the derivability relation of the metatheory ($\vdash_{\mathrm{M}}$) under the reflection rules. It is worth noticing that we must keep distinct the formulas of the object theory from those of the meta theory. We therefore assume that formulas are implicitly labeled by the language they belong to, and that the inference rules Δ_O and in Δ_M are applicable only to formulas in the corresponding language. In most of the previous work (see for instance [17, 31]), the labeling is explicit. Thus, for instance, conjunction introduction in Δ_O, conjunction introduction in Δ_M, and Rup would be written, respectively, as

$$\frac{O:A \quad O:B}{O:A\wedge B} \qquad \frac{M:A \quad M:B}{M:A\wedge B} \qquad \frac{O:A}{M:\bullet(\text{``}A\text{''})}$$

Since we are considering only two theories, the context will always make clear the language a formula belongs to. Formulas containing $\bullet$ are assumed to be in L_M, while formulas with no $\bullet$ are assumed to be in L_O, unless explicitly stated to the contrary.

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 $\Gamma \vdash_{\mathrm{OM}} A$, 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 $\vdash_{\mathrm{OM}} A$, if it is derivable from the empty set. Examples of deductions in OM pairs are the following.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\bullet(\text{``}A\text{''})^{(1)}}{A}\,\text{Rdw} \qquad \dfrac{\bullet(\text{``}A\supset B\text{''})^{(2)}}{A\supset B}\,\text{Rdw}}{B}\,\supset\text{E}}{\bullet(\text{``}B\text{''})}\,\text{Rup}_{\text{r}}}{\bullet(\text{``}A\text{''})\supset\bullet(\text{``}B\text{''})}\,\supset\text{I disch (1)}}{\bullet(\text{``}A\supset B\text{''})\supset(\bullet(\text{``}A\text{''})\supset\bullet(\text{``}B\text{''}))}\,\supset\text{I disch (2)} \tag{2}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A^{(1)}}{\bullet(\text{``}A\text{''})}\,\text{Rup} \qquad \neg\bullet(\text{``}A\text{''})^{(2)}}{\bot}\,\supset\text{E}}{\bullet(\text{``}\bot\text{''})}\,\bot}{\bot}\,\text{Rdw}}{\neg A}\,\supset\text{I disch (1)}}{\bullet(\text{``}\neg A\text{''})}\,\text{Rup}}{\neg\bullet(\text{``}A\text{''})\supset\bullet(\text{``}\neg A\text{''})}\,\supset\text{I disch (2)} \tag{3}$$

$$\dfrac{\dfrac{\bot^{(1)}}{\neg\bot}\,\supset\text{I disch (1)}}{\neg\bullet(\text{``}\bot\text{''})}\,\text{Rup}_{\text{r}}^{\text{n}} \tag{4}$$

$\mathrm{TH}_{\mathrm{OM}}(O) \subseteq L_O$ 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 $\mathrm{TH}(M) \subseteq \mathrm{TH}_{\mathrm{OM}}(M)$. 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 *meta theory 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')=$\mathrm{TH}_{\mathrm{OM}}(M)$ (see again Figure 1).

2.3 Meaningful OM pairs

The main objective of this paper is to characterize the object theory and the metatheory generated by any combinations of reflection rules. To keep the presentation reasonably short, we consider only some combinations of meaningful reflection rules. The selection criteria of the group of reflection rules to be analyzed is the following.

- We first concentrate ourself on the positive rules (mainly used in the literature). The duality results in Section 5 allow us to extend the results to OM pairs mainly involving negative rules.
- We do not consider the combinations of reflection rules that go all in the same direction (either from O to M or in the opposite direction). For instance we do not consider $\mathrm{Rup}+\mathrm{Rup}_r^n$. The results about these combinations can be obtained by composing the results of the single reflection rules.
- We do not consider the combinations of reflection rules which are deductively equivalent to some other combination already considered. For instance we do not consider $\mathrm{Rup}+\mathrm{Rdw}+\mathrm{Rdw}^n$ as it is deductively equivalent to $\mathrm{Rup}+\mathrm{Rdw}$. (It can be easily shown that Rdw^n is a derived inference rule when we have $\mathrm{Rup}+\mathrm{Rdw}$.)
- Finally we do not consider those combinations of reflection rules in which the reflection down rules are "more restricted" than the reflection up rules. In other words we require that if there is a restricted reflection down then all the reflection up rules must be restricted. For instance we do not consider the combination $\mathrm{Rdw}_r+\mathrm{Rup}$ or the combination $\mathrm{Rdw}+\mathrm{Rdw}_r^n+\mathrm{Rup}$. The reasons for this choice are pragmatical: to our knowledge these combinations have received little or no interest in the past.

The remaining combinations of reflection rules are:

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

In each section we first concentrate on the combinations labeled with *. At the end of each section we describe how the results can be extended to the other combinations.

3 A relative characterization

A first characterization of the different combinations of reflection rules can be obtained by comparing them. Any pair of combinations of reflection rules is compared according to the following three criteria:

1. the provability relation of the object theory, namely $\mathrm{TH}_{\mathrm{OM}}(O)$,
2. the provability relation of the meta theory, namely $\mathrm{TH}_{\mathrm{OM}}(M)$, and

3. the OM derivability relation $\vdash_{\mathrm{OM}}$.

The main result of this section is a partial order of the combinations of reflection rules. This partial order is a lattice with top element $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup}_{\mathrm{r}}^{\mathrm{n}}$, the strongest combination, and a bottom element $\emptyset$, the weakest combination.

Definition 3 For each two OM pairs OM_1 and OM_2,

1. $\mathrm{OM}_1 \leq_O \mathrm{OM}_2$ if $\mathrm{TH}_{\mathrm{OM}_1}(O) \subseteq \mathrm{TH}_{\mathrm{OM}_2}(O)$;
2. $\mathrm{OM}_1 \leq_M \mathrm{OM}_2$ if $\mathrm{TH}_{\mathrm{OM}_1}(M) \subseteq \mathrm{TH}_{\mathrm{OM}_2}(M)$;
3. $\mathrm{OM}_1 \leq_D \mathrm{OM}_2$ if $\vdash_{\mathrm{OM}_1} \subseteq \vdash_{\mathrm{OM}_2}$;
4. $\mathrm{OM}_1 =_X \mathrm{OM}_2$ if $\mathrm{OM}_1 \leq_X \mathrm{OM}_2$ and $\mathrm{OM}_2 \leq_X \mathrm{OM}_1$, 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 Let OM(RR) be an OM pair composed of an object theory and a meta theory connected by the set of reflection rules (RR). For any combination of reflection rules $(\mathrm{RR})_1$, $(\mathrm{RR})_2$:

1. $(\mathrm{RR})_1 \leq_O (\mathrm{RR})_2$ if, for any $\mathrm{OM}(\mathrm{RR})_1$ and $\mathrm{OM}(\mathrm{RR})_2$, with the same objec theory and meta theory,$\mathrm{OM}(\mathrm{RR})_1 \leq_O \mathrm{OM}(\mathrm{RR})_2$;
2. $(\mathrm{RR})_1 \leq_M (\mathrm{RR})_2$ if, for any $\mathrm{OM}(\mathrm{RR})_1$ and $\mathrm{OM}(\mathrm{RR})_2$ with the same objec theory and meta theory, $\mathrm{OM}(\mathrm{RR})_1 \leq_M \mathrm{OM}(\mathrm{RR})_2$;
3. $(\mathrm{RR})_1 \leq_D (\mathrm{RR})_2$ if, for any $\mathrm{OM}(\mathrm{RR})_1$ and $\mathrm{OM}(\mathrm{RR})_2$ with the same objec theory and meta theory, $\mathrm{OM}(\mathrm{RR})_1 \leq_D \mathrm{OM}(\mathrm{RR})_2$;
4. For any $X \in \{O, M, D\}$, $(\mathrm{RR})_1 =_X (\mathrm{RR})_2$, if both $(\mathrm{RR})_1 \leq_X (\mathrm{RR})_2$ and $(\mathrm{RR})_2 \leq_X (\mathrm{RR})_1$.

The orders defined above are not completely independent of one another. For instance, if two combinations of reflection rules are equivalent (i.e., $(\mathrm{RR})_1 =_D (\mathrm{RR})_2$) then they generate the same object theory and the same meta theory (i.e., $(\mathrm{RR})_1 =_O (\mathrm{RR})_2$ and $(\mathrm{RR})_1 =_M (\mathrm{RR})_2$). The relations among the three orders are summarized in Figure 2. In this figure, an arrow connects X to Y, if and only if $(\mathrm{RR})_1 X (\mathrm{RR})_2$ implies $(\mathrm{RR})_1 Y (\mathrm{RR})_2$. Thus, for instance, Figure 2 states that, if $(\mathrm{RR})_1 \leq_D (\mathrm{RR})_2$, then $(\mathrm{RR})_1 \leq_O (\mathrm{RR})_2$ and $(\mathrm{RR})_1 \leq_M (\mathrm{RR})_2$.

The relations among sets of reflection rules are represented in Figure 3.

Theorem 5 The graph in Figure 3 is *correct.* If there is an arc labelled with X from $(\mathrm{RR})_1$ to $(\mathrm{RR})_2$, then $(\mathrm{RR})_1 X (\mathrm{RR})_2$.

Proof We prove the correctness of each arc and label in Figure 3.

1. All the instances of the relation $\leq_D$ stated in the graph, with the exception of $\mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}} \leq_D \mathrm{Rup} + \mathrm{Rdw}$, are consequences of Proposition 6. This special case is proved in Proposition 7 item (1).

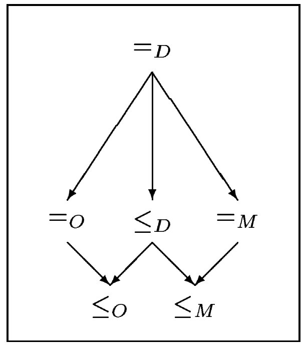

Figure 2: Relations among orders.

2. $\emptyset =_O \mathrm{Rup_r} =_O \mathrm{Rup}$. $\emptyset \leq_O \mathrm{Rup_r} \leq_O \mathrm{Rup}$ is a consequence of item (1) of this proof. Viceversa $\mathrm{Rup} \leq_O \mathrm{Rup_r} \leq_O \emptyset$ holds since in OM pairs with only reflection up all the proofs of object theorems do not contain applications of reflection rules.

3. $\emptyset =_M \mathrm{Rdw_r} =_M \mathrm{Rdw}$. The proof is analogous to that of item (2).

4. $\mathrm{Rup_r} =_M \mathrm{Rup}$. This statement holds since in the OM pairs with a single reflection up, the premises of any application of a reflection up in a proof (with no undischarged assumptions) cannot depend from any assumption on O. In fact there is no reflection down that allows to switch back into O to discharge such assumptions.

5. $\mathrm{Rdw_r} =_O \mathrm{Rdw}$. The proof is analogous to that of item (4).

6. $\mathrm{Rup_r} + \mathrm{Rdw} =_O \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$. $\mathrm{Rup_r} + \mathrm{Rdw} \leq_O \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$ is a consequence of item (1). For the proof of $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} \leq_O \mathrm{Rup_r} + \mathrm{Rdw}$ see Proposition 8.

7. $\mathrm{Rup_r^n} + \mathrm{Rdw} =_M \mathrm{Rup^n} + \mathrm{Rdw}$. $\mathrm{Rup_r^n} + \mathrm{Rdw} \leq_M \mathrm{Rup^n} + \mathrm{Rdw}$ is a consequence of item (1). The converse is proved in Proposition 9.

8. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} =_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} \leq_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$; is a consequence of Item 1. The converse is a consequence of Proposition 9.

9. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n} =_O \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup^n}$ is a consequence of item (8) of this proof and of Proposition 10. Indeed, by item (8), we have that the two sets of reflection rules generate the same metatheory. Furthermore, they both contain $\mathrm{Rup_r}$ and $\mathrm{Rdw_r}$. As a consequence, by Proposition 10, we can infer that they generate the same object theory.

10. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n} =_M \mathrm{Rup} + \mathrm{Rdw}$. $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n} \leq_M \mathrm{Rup} + \mathrm{Rdw}$ is a consequence of Proposition 7 item (1). $\mathrm{Rup} + \mathrm{Rdw} \leq_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n}$ is proved in Proposition 11.

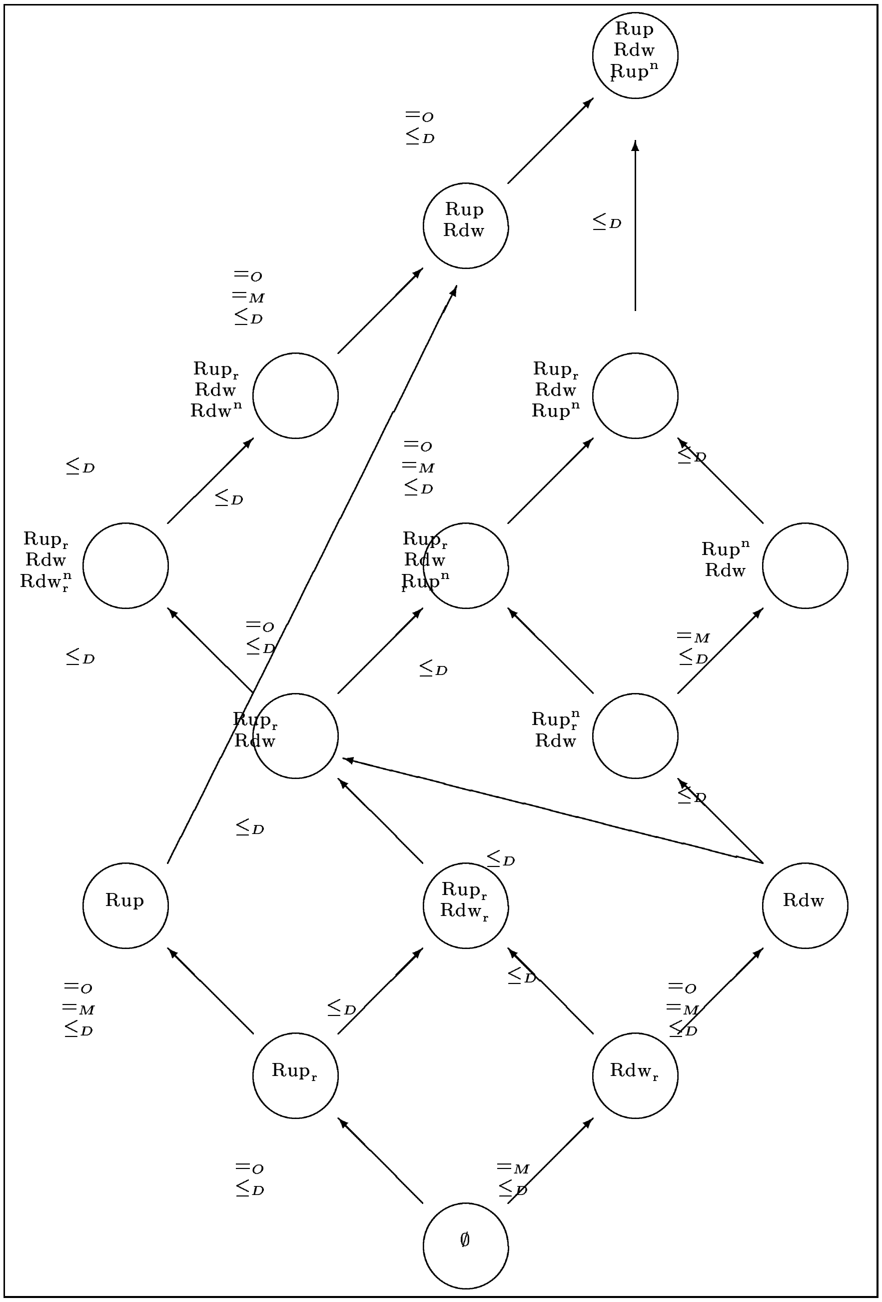

Figure 3: Partial orders of (RR)'s.

11. $\text{Rup}_\text{r} + \text{Rdw} + \text{Rdw}^\text{n} =_O \text{Rup} + \text{Rdw}$ can be proved from item (10) and Proposition 10 and by reasoning as in item (9).

12. $\text{Rup} + \text{Rdw} =_O \text{Rup} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$ can be proved by reasoning as in Proposition 8.

Proposition 6 For any two sets of reflection rules $(\text{RR})_1$ and $(\text{RR})_2$, any of the following conditions implies that $(\text{RR})_1 \leq_D (\text{RR})_2$

1. $(\text{RR})_1 \subseteq (\text{RR})_2$;
2. $(\text{RR})_1$ is obtained by restricting some reflection rules of $(\text{RR})_2$;
3. For any $r \in (\text{RR})_1$, $r \leq_D (\text{RR})_2$.

Proposition 7

1. $\text{Rdw}^\text{n} \leq_D \text{Rup} + \text{Rdw}$;
2. $\text{Rup}^\text{n} \leq_D \text{Rup} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$;
3. $(\text{RR}) \leq_D \text{Rup} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$, for any set of reflection rules (RR).

Proof To prove item (1), we show that Rdw^n is a derived reflection rule in Rup + Rdw, namely, we provide a deduction of $\neg A$ from $\neg \bullet (``A")$ using Rup + Rdw.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A^{(1)}}{\bullet(``A")}\,\text{Rup} \qquad \neg\bullet(``A")}{\bot}\,\supset\!\text{E}_M}{\bullet(``\bot")}\,\bot_M}{\bot}\,\text{Rdw}}{\neg A}\,\supset\!\text{I}_O\ \text{disch. (1)}$$

To prove item (2) we show that Rup^n is a derived reflection rule in $\text{Rup}+\text{Rdw}+\text{Rup}_\text{r}^\text{n}$, namely, we provide a deduction of $\neg \bullet (``A")$ form $\neg A$ using $\text{Rup} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\bullet(``A")^{(1)}}{A}\,\text{Rdw} \qquad \neg A}{\bot}\,\supset\!\text{E}_O}{\bullet(``\bot")}\,\text{Rup} \qquad \dfrac{\dfrac{\bot^{(2)}}{\neg\bot}\,\supset\!\text{I}_O\ \text{disch. (2)}}{\neg\bullet(``\bot")}\,\text{Rup}_\text{r}^\text{n}}{\bot}\,\supset\!\text{E}_M}{\neg\bullet(``A")}\,\supset\!\text{I}_M\ \text{disch. (1)}$$

Item (3) is a consequence of items (1), (2) and Proposition 6.

Proposition 8 $\text{Rup}_\text{r} + \text{Rdw} + \text{Rup}_\text{r}^\text{n} \leq_O \text{Rup}_\text{r} + \text{Rdw}$.

Proof Let MK and MC be two OM pairs composed of O and M connected by the reflection rules $\text{Rup}_\text{r} + \text{Rdw}$ and $\text{Rup}_\text{r} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$ respectively. For each L_O-wff A_O and for each L_M-wff A_M we have that:

$$\vdash_{\text{MC}} A_M \implies \neg\bullet(``\bot") \vdash_{\text{MK}} A_M \tag{5}$$

$$\vdash_{\text{MC}} A_O \implies \vdash_{\text{MK}} A_O \tag{6}$$

We start by proving (5). Let Π be a proof of a formula A_M of the meta theory of MC containing an application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. We transform Π into a deduction Π' as follows:

$$\begin{array}{c}\Pi_1\\ \dfrac{\neg B}{\neg\bullet(\text{``}B\text{''})}\,\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\\ \Pi_2\\ A_M\end{array} \quad\Longrightarrow\quad \begin{array}{c}\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}\Pi_1\\ \neg B\end{array}\quad \dfrac{\bullet(\text{``}B\text{''})^{(1)}}{B}\,\mathrm{Rdw}}{\bot}\,\supset\mathrm{E}_O}{\bullet(\text{``}\bot\text{''})}\,\mathrm{Rup}_{\mathrm{r}}\qquad \neg\bullet(\text{``}\bot\text{''})}{\bot}\,\supset\mathrm{E}_M}{\neg\bullet(\text{``}B\text{''})}\,\supset\mathrm{I}\ \text{disch. (1)}\\ \Pi_2\\ A_M\end{array} \qquad (7)$$

Transformation (7) replaces the applications of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ by the assumption $\neg\bullet(\text{``}\bot\text{''})$. Any assumption discharged in the deduction on the left side of (7) is also discharged in the deduction on the right side of (7). By applying (7) to all the occurrences of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ of $\Pi$, we obtain a deduction of $A_M$ from $\neg\bullet(\text{``}\bot\text{''})$ which does not contain any application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. This is therefore a deduction of A_M from $\bullet(\text{``}\bot\text{''})$ in MK.

Property (6) is proved as follows. Let Π be a proof of A_O in MC. If Π contains no applications of Rdw, then it contains no applications of reflection rules, and $A_O \in \mathrm{TH}(O)$. Hence A_O is provable in MK too. If it contains an application of Rdw, then Π must be of the form:

$$\begin{array}{c}\begin{array}{c}\Pi_1\\ \dfrac{\bullet(\text{``}A_1\text{''})}{A_1}\,\mathrm{Rdw}\end{array}\quad \ldots \quad \begin{array}{c}\Pi_n\\ \dfrac{\bullet(\text{``}A_n\text{''})}{A_n}\,\mathrm{Rdw}\end{array}\\ \Pi_O\\ A_O\end{array}$$

for some $n \geq 1$, such that Π_O is a deduction of A_O in O from $A_1, \ldots, A_n$, and each Π_i is a proof of $\bullet(\text{``}A_i\text{''})$ in $MC$. By Property (5) we have that, for each $1 \leq i \leq n$, $\neg\bullet(\text{``}\bot\text{''}) \vdash_{\mathrm{MK}} \bullet(\text{``}A_i\text{''})$. The following deduction shows that each A_i is a theorem of MK.

$$\dfrac{\dfrac{\bullet(\text{``}\bot\text{''}) \vee \neg\bullet(\text{``}\bot\text{''}) \qquad \dfrac{\dfrac{\dfrac{\bullet(\text{``}\bot\text{''})^{(1)}}{\bot}\,\mathrm{Rdw}}{A}\,\bot_O}{\bullet(\text{``}A_i\text{''})}\,\mathrm{Rup}_{\mathrm{r}} \qquad \begin{array}{c}\neg\bullet(\text{``}\bot\text{''})^{(2)}\\ \vdots\ \text{prop. (5)}\\ \bullet(\text{``}A_i\text{''})\end{array}}{\bullet(\text{``}A_i\text{''})}\,\vee\mathrm{E}_M\ \text{disch. (1) and (2)}}{A_i}\,\mathrm{Rdw}$$

From the fact that all A_i's are theorems of the object theory of MK and that A_O is derivable from $A_1, \ldots, A_n$, we conclude that A_O is provable in the object theory of MK.

Proposition 9 $\mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw} \leq_M \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}$.

Proof We show how to replace every application of $\mathrm{Rup}^{\mathrm{n}}$ with an application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ in the proofs of meta theorems. Let A_M be a meta theorem and let Π be a proof A_M. Suppose that Π contains an application of $\mathrm{Rup}^{\mathrm{n}}$ which is not an application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ (i.e., the premise depends on a not empty set of assumptions Γ_O in the object theory). Since in Π all the assumptions are discharged, there must be at least an application of Rdw occurring below the application of $\mathrm{Rup}^{\mathrm{n}}$. This very last application is necessary in order to discharge the assumptions Γ_O (remember that the conclusion A_M does not depend on any assumptions). Π is therefore of the following form.

$$
\begin{array}{ll}
\begin{array}{c}
\Gamma_O \\
\Pi_1 \\
\dfrac{\neg B}{\neg \bullet(\text{“}B\text{”})}\ \mathrm{Rup}^{\mathrm{n}}
\end{array} & \longleftarrow \text{ Last application of } \mathrm{Rup}^{\mathrm{n}} \text{ which is not an application of } \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} \\
\begin{array}{c}
\Pi_2 \\
\dfrac{\bullet(\text{“}C\text{”})}{C}\ \mathrm{Rdw}
\end{array} & \longleftarrow \text{ The first application of Rdw after } \mathrm{Rup}^{\mathrm{n}} \\
\begin{array}{c}
\Pi_3 \\
\dfrac{\neg D}{\neg \bullet(\text{“}D\text{”})}\ \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}
\end{array} & \begin{array}{l} \longleftarrow \text{ All assumptions in } \Gamma_O \text{ are discharged} \\ \longleftarrow \text{ Last application of } \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} \end{array} \\
\begin{array}{c}
\Pi_4 \\
A_M
\end{array} &
\end{array}
\tag{8}
$$

This application of $\mathrm{Rup}^{\mathrm{n}}$ can be removed from Π by rewriting Π as follows:

$$
\begin{array}{c}
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}\Gamma_O \\ \Pi_1 \\ \neg B\end{array} \quad \dfrac{\bullet(\text{“}B\text{”})^{(1)}}{B}\ \mathrm{Rdw}}{\bot}\ \supset\mathrm{E}}{C}\ \bot}{\begin{array}{c}\Pi_3 \\ \neg D^{(*)}\end{array}}}{\neg \bullet(\text{“}D\text{”})}\ \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} \quad \bullet(\text{“}D\text{”})^{(2)}}{\bot}\ \supset\mathrm{E}}{\neg \bullet(\text{“}B\text{”})}\ \supset\mathrm{I}\ \text{disch. (1)}}{\begin{array}{c}\Pi_2 \\ \bullet(\text{“}C\text{”})\end{array}}}{C}\ \mathrm{Rdw}}{\begin{array}{c}\Pi_3 \\ \neg D^{(**)}\end{array}}}{\neg \bullet(\text{“}D\text{”})}\ \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} \quad \bullet(\text{“}D\text{”})^{(2)}}{\dfrac{\bot}{\neg \bullet(\text{“}D\text{”})}\ \supset\mathrm{I}\ \text{disch. (2)}}\ \supset\mathrm{E} \\
\Pi_4 \\
A_M
\end{array}
\tag{9}
$$

In Deduction (9) the assumptions in Γ_O are discharged before the occurrence of $\neg D$ labelled with $^{(*)}$. This ensures the applicability of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ to this occurrence of $\neg D$. The other application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ in (9) to the occurrence of $\neg D$ labelled with $^{(**)}$ is also allowed. Indeed $\neg D^{(**)}$ in (9) depends on $\bullet(\text{“}D\text{”})$ plus the assumptions on which $\neg \bullet(\text{“}D\text{”})$ depends in (8). Any other assumption in (8) is discharged in (9). Indeed an assumption of (8) is made in a Π_i and discharged in a Π_j, with $i \leq j$ and $i, j = 1, 2, 3, 4$. Deduction (9) maintains the same structure; that is, it is built in such a way that, for any $i \leq j$, Π_j occurs below Π_i. Finally the assumption $\bullet(\text{“}D\text{”})$ is discharged by an application of $\supset$I.

Proposition 10 For any pair of sets of reflection rules $(\mathrm{RR})_1$ and $(\mathrm{RR})_2$, if $\mathrm{Rup_r} \leq_D (\mathrm{RR})_1$ and $\mathrm{Rdw_r} \leq_D (\mathrm{RR})_2$, then $(\mathrm{RR})_1 \leq_M (\mathrm{RR})_2$ implies $(\mathrm{RR})_1 \leq_O (\mathrm{RR})_2$.

Proof Let OM_1 and OM_2 be two OM pairs composed of O and M connected by $(\mathrm{RR})_1$ and $(\mathrm{RR})_2$, respectively. If A is a theorem of the object theory of OM_1, then, by $\mathrm{Rup_r}$, $\bullet(\text{“}A\text{”})$ is a theorem of the metatheory of OM_1. The hypothesis $\mathrm{OM}_1 \leq_M \mathrm{OM}_2$ implies that $\bullet(\text{“}A\text{”})$ is a theorem also of the metatheory of OM_2. Therefore A is provable (with an application of $\mathrm{Rdw_r}$) in the object theory of OM_2.

Proposition 11 $\mathrm{Rup} + \mathrm{Rdw} \leq_M \mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw^n}$;

Proof Let OM_1 and OM_2 be two OM pairs composed of O and M connected by $\mathrm{Rup}+\mathrm{Rdw}$ and $\mathrm{Rup_r}+\mathrm{Rdw}+\mathrm{Rdw^n}$, respectively. We provide a method for transforming a proof Π of A_M in OM_1 into a proof Π' of A_M in OM_2. This is done by replacing any assumption of Π of a wff of the object language, with suitable assumptions of wffs of the meta language and applications of $\mathrm{Rdw^n}$.

Let C be an assumption of a wff in the object language. Since Π is a proof, C must be discharged. If C is discharged by $\supset$I, then we apply Transformation (10) to Π. If C is discharged by $\bot_c$, then we apply Transformation (11).

$$\begin{array}{c} C \\ \Pi_1 \\ \dfrac{B}{C \supset B}\supset\mathrm{I} \\ \Pi_2 \\ A_M \end{array} \Longrightarrow \dfrac{\begin{array}{c}\vdots\\ \bullet(\text{“}C\text{”}) \vee \neg\bullet(\text{“}C\text{”})\end{array} \quad \begin{array}{c} \dfrac{\bullet(\text{“}C\text{”})}{C}\mathrm{Rdw} \\ \Pi_1 \\ \dfrac{B}{C\supset B}\supset\mathrm{I} \\ \Pi_2 \\ A_M\end{array} \quad \begin{array}{c} \dfrac{\dfrac{\neg\bullet(\text{“}C\text{”})}{\neg C}\mathrm{Rdw^n}}{C\supset B} \\ \Pi_2 \\ A_M \end{array}}{A_M}\vee\mathrm{E} \tag{10}$$

$$\begin{array}{c} \neg C \\ \Pi_1 \\ \dfrac{\bot}{C}\bot_c \\ \Pi_2 \\ A \end{array} \Longrightarrow \dfrac{\begin{array}{c}\vdots\\ \bullet(\text{“}C\text{”}) \vee \neg\bullet(\text{“}C\text{”})\end{array} \quad \begin{array}{c} \dfrac{\neg\bullet(\text{“}C\text{”})}{\neg C}\mathrm{Rdw^n} \\ \Pi_1 \\ \dfrac{\bot}{C}\bot_c \\ \Pi_2 \\ A\end{array} \quad \begin{array}{c} \dfrac{\bullet(\text{“}C\text{”})}{C}\mathrm{Rdw} \\ \Pi_2 \\ A \end{array}}{A}\vee\mathrm{E} \tag{11}$$

Transformations (10) and (11) do not introduce undischarged assumptions. Therefore they transform a proof of A_M into another proof of A_M. By repeatedly applying (10) and (11) for each assumption in the object language we obtain a proof Π' of A_M where all the assumptions are wffs of the meta language. This implies that all the applications of Rup in Π' are applications of $\mathrm{Rup_r}$, and, therefore, that Π' is a proof of A_M in OM_2.

Table 1 summarizes the relations among all the sets of reflection rules. It represents relations of the form $(\mathrm{RR})_1 X (\mathrm{RR})_2$, where, for each square $(\mathrm{RR})_1$ is the set of reflection rules in the left hand side of the table, X is a symbol inside the square, and $(\mathrm{RR})_2$ is the set of reflection rules in the top row. The facts represented in this table are derivable by transitive closure on the relations stated in the graph of figure 3. The relations $\leq_O$,$\leq_M$ and $\leq_D$ are indeed partial orders, and $=_O$, $=_M$, and $=_D$ are equivalence relations.

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

Table 1: Relations among (RR)'s.

We draw some final remarks on the relations proved in this section. A first remark concerns the rule Rup^{n}. When $\bullet(``A")$ can be interpreted as "A is provable in the object theory", (i.e., when the object theory and the metatheory are connected by $\text{Rup}_{\text{r}} + \text{Rdw}$) Rup^{n} corresponds to the assertion of the consistency of the object theory. This correspondence is formally supported, in the proof of Proposition 8, by proving that any application of $\text{Rup}_{\text{r}}^{\text{n}}$ can be replaced by the assumption $\neg \bullet (``\bot")$ (see transformation (11)). Notice however that, differently to what happens in other cases where the object theory results to be modified by this assertion, Proposition 8 shows that the assertion of the consistency of the object theory, either by adding Rup^{n} or by assuming $\neg \bullet (``\bot")$, does not affect the object theory itself. In our framework, therefore, one can assert the consistency of the object theory in a pure descriptive way. Furthermore notice that, if the metatheory asserts the consistency of the object theory when the object theory is actually inconsistent, than the meta theory will result inconsistent too.

A second remark concerns the asymmetry between the relation $\text{Rup}_{\text{r}} + \text{Rdw} \leq_M \text{Rup} + \text{Rdw}$ and $\text{Rup}_{\text{r}}^{\text{n}} + \text{Rdw} =_M \text{Rup}^{\text{n}} + \text{Rdw}$. In the first case, indeed, relaxing the restriction on the (positive) reflection up strengthens the meta theory; in the second case, instead, such a relaxation leaves the metatheory unchanged.

4 An axiomatic characterization

Consider Figure 1 in Section 1, our goal in this section is to *axiomatically characterize* the effects of the reflection rules on the theorems of O and M. This means that we are looking for an axiomatic extension of O and M, to $O' = \langle L_O, \Omega'_O, \Delta_O \rangle$ and $M' = \langle L_M, \Omega'_M, \Delta_M \rangle$, such that the set of theorems of O' and M" are equal $\text{TH}_{\text{OM}}(O)$ and $\text{TH}_{\text{OM}}(M)$, respectively. Our first step is to define Ω'_O and Ω'_M in terms of simple and commonly used axiom schemata, (Subsection 4.1). However, this approach seems rather ad hoc and, in any case, it does not allow us to characterize all the sets of reflection rules. We then move to a more uniform approach and characterize Ω'_O and Ω'_M via a fixpoint equation (Subsection 4.2). We conclude by comparing the two approaches and by formally proving that the fixpoint analysis allows us to characterize combinations of reflection rules not characterizable axiomatially (Subsection 4.3). Most interestingly, we show that one such combination is $\text{Rup}_{\text{r}} + \text{Rdw}_{\text{r}}$, that is, the most used combination of reflection rules which can be found in the literature.

We 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 $\Gamma \subseteq L_O$, $\bullet(``\Gamma")$ denotes the set of wffs $\{\bullet(``A") : A \in \Gamma\} \subseteq L_M$. For any set of wffs $\Gamma \subseteq L_M$, $\bullet^{-1}(``\Gamma")$ denotes the set of wffs $\{A : \bullet(``A") \in \Gamma\} \subseteq L_O$.

4.1 Characterization via simple axiom schemata

A first characterization of reflection rules is given in terms of the most common meta axiom schemata, which have been extensively used in the literature. Examples of such meta axioms are the following:

$$\bullet(``A \supset B") \supset (\bullet(``A") \supset \bullet(``B")) \tag{K}$$

$$\neg \bullet (``A") \supset \bullet(``\neg A") \tag{Comp}$$

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

where A, B must be understood as schematic variables (parameters) ranging over L_O. Axioms of this kind, and in particular the axioms listed above, usually have an intuitive meaning and define (what in our terminology is) the global meta property captured by the corresponding metatheory. Thus (K) states that the interpretation of $\bullet$ is closed under *modus ponens*. This is a necessary (but not sufficient) condition for $\bullet$ to be a provability predicate. The axiom (Comp) states that each interpretation of $\bullet$ is saturated, meaning that for each $A \in L_0$, either A or its negation belongs to the interpretation of $\bullet$. This is a necessary (but not sufficient) condition for $\bullet$ to be a truth predicate. The axiom (nTbot) states that $\bot$ does not belong to the interpretation of $\bullet$. This is a necessary (but not sufficient) condition for a set of formulas to be consistent. In the following we show how most of the effects of the reflection rules can be characterized by suitable combination of the above axiom schemata.

Theorem 12 Let OM be an OM pair composed of O and M connected by a set of reflection rules (RR). If (RR) is Rdw_r or Rdw, then:

$$\begin{aligned} \text{TH}_{\text{OM}}(O) &= \text{TH}(O + \bullet^{-1}(\text{“TH}(M)\text{”})) &(12)\\ \text{TH}_{\text{OM}}(M) &= \text{TH}(M) &(13) \end{aligned}$$

If (RR) is Rup_r or Rup, then:

$$\begin{aligned} \text{TH}_{\text{OM}}(O) &= \text{TH}(O) &(14)\\ \text{TH}_{\text{OM}}(M) &= \text{TH}(M + \bullet(\text{“TH}(O)\text{”})) &(15) \end{aligned}$$

Proof (Theorem 12) By Table 1, we have that the metatheory generated by Rdw and Rdw_r is the same as that generated by the empty set of reflection rules, which is $\text{TH}(M)$. This proves (13). To prove (12) we consider only Rdw_r as, from Table 1, Rdw_r and Rdw generate the same object theory. Notice that the premise of each application of Rdw_r in a proof of an object theorem A, must be provable in the metatheory of OM, and, by (13), in M itself. This implies that, given a proof Π of A in OM, the proof of $\bullet(\text{“}B\text{”})$ occurring above any application of Rdw_r in Π, can be replaced by B, which belongs to $\bullet^{-1}(\text{“TH}(M)\text{”})$. The result is a proof of A in $O{+}\bullet^{-1}(\text{“TH}(M)\text{”})$. This proves that $\text{TH}(O{+}\bullet^{-1}(\text{“TH}(M)\text{”})) \subseteq \text{TH}_{\text{OM}}(O)$. Viceversa, to prove $\text{TH}_{\text{OM}}(O) \subseteq \text{TH}(O + \bullet^{-1}(\text{“TH}(M)\text{”}))$, it is enough to observe that, by Rdw_r, $\bullet^{-1}(\text{“TH}(M)\text{”}) \subseteq \text{TH}_{\text{OM}}(O)$. (14) and (15) can be proved analogously.

Theorem 13 Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR). Then:

$$\begin{aligned} \text{TH}_{\text{OM}}(O) &= \text{TH}(O + \bullet^{-1}(\text{“TH}_{\text{OM}}(M)\text{”}))\\ \text{TH}_{\text{OM}}(M) &= \text{TH}(M + \bullet(\text{“TH}(O)\text{”}) + \Gamma) \end{aligned}$$

where:

1. If (RR) is $\text{Rup}_\text{r} + \text{Rdw}$, then $\Gamma = (K)$;
2. If (RR) is $\text{Rup}_\text{r} + \text{Rdw} + \text{Rup}_\text{r}^\text{n}$, then $\Gamma = (K) \cup (\text{nTbot})$;

3. If (RR) is Rup + Rdw, then $\Gamma = (\mathrm{K}) \cup (\mathrm{Comp})$;
4. If (RR) is Rup + Rdw + $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ then $\Gamma = (\mathrm{K}) \cup (\mathrm{Comp}) \cup (\mathrm{nTbot})$.

Theorem 13 tells us that, for all the combinations of reflection urles listed in items 1–4, there is a schematic characterization of the meta theory in terms of the axioms (K), (nTbot) and (Comp). Notice that, Table 1 states that $\mathrm{Rup} + \mathrm{Rdw} =_M \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$ and that $\mathrm{Rup} + \mathrm{Rdw} =_O \mathrm{Rup}_{\mathrm{r}} + \mathrm{Rdw} + \mathrm{Rdw}^{\mathrm{n}}$; we can therefore extend the results presented in Theorem 13 for Rup+Rdw also to $\mathrm{Rup}_{\mathrm{r}}+\mathrm{Rdw}+\mathrm{Rdw}^{\mathrm{n}}$.

Proof (Theorem 13) Let us start by showing that

$$\mathrm{TH}_{\mathrm{OM}}(O) \subseteq \mathrm{TH}(O + \bullet^{-1}(``\mathrm{TH}_{\mathrm{OM}}(M)"))$$

If $A \in \mathrm{TH}_{\mathrm{OM}}(O)$, then, by $\mathrm{Rup}_{\mathrm{r}}$, we have that $\bullet(``A") \in \mathrm{TH}_{\mathrm{OM}}(M)$ and, therefore, $A \in \bullet^{-1}(``\mathrm{TH}_{\mathrm{OM}}(M)") \subseteq \mathrm{TH}(O + \bullet^{-1}(``\mathrm{TH}_{\mathrm{OM}}(M)"))$. Viceversa, it is enough to observe that $\bullet^{-1}(``\mathrm{TH}_{\mathrm{OM}}(M)") \subseteq \mathrm{TH}_{\mathrm{OM}}(O)$.

Let us prove that $\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + \bullet(``\mathrm{TH}(O)") + \Gamma)$. Notice that, $\mathrm{TH}(M + \bullet(``\mathrm{TH}(O)") + \Gamma) \subseteq \mathrm{TH}_{\mathrm{OM}}(M)$ holds as all the formulas in $\bullet(``\mathrm{TH}(O)")$ and all the formulas in Γ are provable in the OM pair with the corresponding combination (RR) (see deductions in (2)–(4)).

Let us now prove $\mathrm{TH}_{\mathrm{OM}}(M) \subseteq \mathrm{TH}(M + \bullet(``\mathrm{TH}(O)") + \Gamma)$. We do this by showing that each application of a reflection rule in a proof of a meta theorem can be replaced by introducing the corresponding axioms. We proceed as follows. In the first step we show how each application of Rup can be replaced with a suitable combination of $\mathrm{Rup}_{\mathrm{r}}$, Rdw and (Comp). In the second step we show how each application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$ can be replaced with a suitable combination of $\mathrm{Rup}_{\mathrm{r}}$, Rdw and (nTbot). Finally, we show how the applications of $\mathrm{Rup}_{\mathrm{r}}$ and Rdw can be replaced by axiom (K) and the set of axioms $\bullet(``\mathrm{TH}(O)")$. This completes the proof.

Step 1: Rup $\Longrightarrow$ $\mathrm{Rup}_{\mathrm{r}}$ + Rdw + (Comp) Let Π be a proof of a meta theorem A. Suppose that Π contains an application of Rup whose premise depends on an assumption C in the object theory. Since Π is a proof of A, C must be discharged somewhere in Π. C can be discharged only by an application of $\supset$E or $\bot_c$. In these cases Π can be modified as follows:

$$\begin{array}{c} C \\ \Pi_1 \\ \dfrac{B}{C \supset B}\supset I \\ \Pi_2 \\ A \end{array} \quad \Longrightarrow \quad \dfrac{\bullet(``C") \vee \neg\bullet(``C") \qquad \begin{array}{c} \dfrac{\bullet(``C")^{(1)}}{C}\mathrm{Rdw} \\ \Pi_1 \\ \dfrac{B}{C \supset B}\supset I \\ \Pi_2 \\ A \end{array} \qquad \begin{array}{c} \dfrac{\dfrac{\dfrac{\neg\bullet(``C")^{(2)} \quad (\mathrm{Comp})}{\bullet(``\neg C")}\supset E}{\neg C}\mathrm{Rdw}}{C \supset B} \\ \Pi_2 \\ A \end{array}}{A}\vee E \text{ disch } (1,2) \tag{16}$$

$$\begin{array}{c} \neg C \\ \Pi_1 \\ \dfrac{\bot}{C}\bot_c \\ \Pi_2 \\ A \end{array} \quad \Longrightarrow \quad \dfrac{\bullet(``\neg C") \vee \neg\bullet(``\neg C") \qquad \begin{array}{c} \dfrac{\bullet(``\neg C")^{(1)}}{\neg C}\mathrm{Rdw} \\ \Pi_1 \\ \dfrac{\bot}{C}\bot_c \\ \Pi_2 \\ A \end{array} \qquad \begin{array}{c} \dfrac{\dfrac{\dfrac{\neg\bullet(``\neg C")^{(2)} \quad (\mathrm{Comp})}{\bullet(``\neg\neg C")}\supset E}{\neg\neg C}\mathrm{Rdw}}{C} \\ \Pi_2 \\ A \end{array}}{A}\vee E \text{ disch } (1,2) \tag{17}$$

By repeated applications of (16) and (17) we obtain a proof with no applications of Rup.

Step 2: $\text{Rup}_r^n \Longrightarrow \text{Rup}_r + \text{Rdw} + (\text{nTbot})$

$$\begin{array}{c}\Pi\\ \dfrac{\neg B}{\neg\bullet(\text{“}B\text{”})}\,\text{Rup}_r^n\\ \Pi'\end{array} \quad\Longrightarrow\quad \begin{array}{c} \dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}\Pi\\ \neg B\end{array} \quad \dfrac{\bullet(\text{“}B\text{”})^{(1)}}{B}\,\text{Rdw}}{\bot}\supset\text{E}}{\bullet(\text{“}\bot\text{”})}\,\text{Rup}_r \qquad (\text{nTbot})}{\bot}\supset\text{E}}{\neg\bullet(\text{“}B\text{”})}\supset\text{E disch } (1)\\ \Pi'\end{array} \tag{18}$$

Step 3: $\text{Rup}_r + \text{Rdw} \Longrightarrow \bullet(\text{“TH}(O)\text{”}) + (\text{K})$ Let B be the premise of the last application of Rup_r in a proof Π of a meta formula in OM. If Π does not contain any application of Rdw, then B is a theorem of the object theory and therefore $\bullet(\text{“}B\text{”})$ is in $\bullet(\text{“TH}(O)\text{”})$. This occurrence of Rup_r can be removed from Π by deleting the subdeduction above $\bullet(\text{“}B\text{”})$. If Π contains an application of Rdw, let A be the consequence of an application of Rdw such that no application of reflection rules occurs in the thread [2] from A to B. Then Π can be modified as follows:

$$\begin{array}{c}\Pi_1\\ \dfrac{\bullet(\text{“}A\text{”})}{A}\,\text{Rdw}\\ \Pi_2\\ \dfrac{B}{\bullet(\text{“}B\text{”})}\,\text{Rup}_r\\ \Pi_3\end{array} \quad\Longrightarrow\quad \begin{array}{c} \dfrac{\dfrac{\dfrac{\begin{array}{c}A\\ \Pi_2\\ B\end{array}}{A\supset B}\supset\text{I}}{\bullet(\text{“}A\supset B\text{”})}\,\text{Rup}_r \qquad \begin{array}{c}\Pi_1\\ \bullet(\text{“}A\text{”})\end{array} \quad (\text{K})}{\bullet(\text{“}B\text{”})}\\ \Pi_3\end{array} \tag{19}$$

By repeated applications of (19) we can remove all the occurrences of Rup_r and Rdw. Note that, since the application of reflection up is restricted, then the proof in the left does not contain any undischarged assumption in O which is then discharged in Π_2. Therefore, the above transformation introduces no undischarged assumptions.

The combinations $\text{Rup}_r^n + \text{Rdw}$ and $\text{Rup}^n + \text{Rdw}$ require a special treatment. The metatheory generated by these combinations is characterizable by the axiom schema (Cons)

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

for any $n \geq 0$. This axiom schema is less common than those considered above. Its intuitive interpretation however is interesting. Indeed, for any consistent set of object formulas S, $\bullet(\text{“}A\text{”})$ can be read as *"A belongs to S"*. Indeed the axiom (Cons) states that $A_1, \ldots, A_n$ belongs to S, and A is contradictory with $A_1, \ldots, A_n$ (condition $\vdash_{O+A_1,\ldots,A_n} \neg A\}$), then A cannot be in S, due to its consistency. Notice that $\bullet(\text{“}A\text{”})$ cannot be simply read as *"A is consistent with O"* as with this reading we have that $\bullet(\text{“}A\text{”}) \supset \bullet(\text{“}A \wedge A\text{”})$ should be a theorem of the metatheory, but this is not the case.

[2] The notion of thread is defined in [30]. Intuitively a thread is a path between a leaf and a node in a deduction tree.

Theorem 14 Let OM be the OM pair composed of O and M connected by the set of reflection rules (RR). If (RR) $=$ $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}} + \mathrm{Rdw}$ then,

$$\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \bullet^{-1}(\text{“}\mathrm{TH}_{\mathrm{OM}}(M)\text{”})) \tag{20}$$
$$\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + (\mathrm{Cons})) \tag{21}$$

If (RR) $= \mathrm{Rup}^{\mathrm{n}} + \mathrm{Rdw}$ then,

$$\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O + \{\bigvee_{k=1}^{n} A_k \;:\; \bigvee_{k=1}^{n} \bullet(\text{“}A_k\text{”}) \in \mathrm{TH}_{\mathrm{OM}}(M)\}) \tag{22}$$
$$\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + (\mathrm{Cons})) \tag{23}$$

Proof (Theorem 14) To prove (20) we observe that if a proof of an object theorem contains some applications of reflection rules, then it is of the following form:

$$\dfrac{\begin{array}{c}\Pi_1\\ \bullet(\text{“}A_1\text{”})\end{array}}{A_1}\,\mathrm{Rdw} \quad \ldots \quad \dfrac{\begin{array}{c}\Pi_n\\ \bullet(\text{“}A_n\text{”})\end{array}}{A_n}\,\mathrm{Rdw}$$
$$\begin{array}{c}\Pi_O\\ A\end{array}$$

where Π_O is a deduction in O. Notice that each $\bullet(\text{“}A_i\text{”})$ is a meta theorem. Indeed $\bullet(\text{“}A_i\text{”})$ cannot depend on any assumptions in the object theory, because the reflection up is restricted. Furthermore, $\bullet(\text{“}A_i\text{”})$ does not depend on any meta assumptions, because A does not depend on meta assumption, and Π_O does not discharge any meta assumption. As a consequence A is in $\mathrm{TH}(O + \bullet^{-1}(\text{“}\mathrm{TH}_{\mathrm{OM}}(M)\text{”}))$. The opposite direction is trivial.

Let us prove (21). It is easy to see that all the instances of (Cons) are provable via $\mathrm{Rdw} + \mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$. The proof is analogous to the proof of axiom (K). Let us prove that any theorem in $\mathrm{TH}_{\mathrm{OM}}(M)$ can be proved in $M + (\mathrm{Cons})$. As usual we replace the applications of reflection rules with the axiom. Suppose that a deduction Π contains an application of $\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$, it is of the form:

$$\dfrac{\begin{array}{c}\Pi_1\\ \bullet(\text{“}A_1\text{”})\end{array}}{A_1}\,\mathrm{Rdw} \quad \ldots \quad \dfrac{\begin{array}{c}\Pi_n\\ \bullet(\text{“}A_n\text{”})\end{array}}{A_n}\,\mathrm{Rdw}$$
$$\dfrac{\begin{array}{c}\Pi_O\\ \neg A\end{array}}{\neg\bullet(\text{“}A\text{”})}\,\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}$$

where Π_O is a proof of $\neg A$ in $O + A_1, \ldots, A_n$. As a consequence $\bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”}) \supset \neg\bullet(\text{“}A\text{”})$ is an instance of (Cons), and therefore the applications of the reflection rules can be removed as follows:

$$\dfrac{\dfrac{\begin{array}{ccc}\Pi_1 & & \Pi_n\\ \bullet(\text{“}A_1\text{”}) & \ldots & \bullet(\text{“}A_n\text{”})\end{array}}{\bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”})} \quad \bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”}) \supset \neg\bullet(\text{“}A\text{”})}{\neg\bullet(\text{“}A\text{”})}\,\supset\!\mathrm{E}$$

To prove (22) we need some notation: for each set of wffs $\Gamma_M \subseteq L_M$, let $(\Gamma_M)^{\vee}_{O}$ be the set $\{A_1 \vee \ldots \vee A_n : \Gamma_M \vdash_{\mathrm{OM}} \bullet(\text{“}A_1\text{”}) \vee \ldots \vee \bullet(\text{“}A_n\text{”})\}$. We prove a more general fact, i.e. that for each set of L_O wffs $\Gamma_O \cup \{A_O\}$ and for each set of L_M wffs Γ_M,

$$\Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_O \quad \Longrightarrow \quad \Gamma_O, (\Gamma_M)_O^{\vee} \vdash_o A_O \tag{24}$$

We proceed by induction on the complexity of a deduction Π of A_O from Γ_O, Γ_M. The cases in which Π is an assumption, an axiom or it ends with an application of a classical rule are trivial. Suppose that Π ends with an application of Rdw. Then Π is of the form:

$$\dfrac{\dfrac{\dfrac{\Pi_1}{\neg A_1}}{\neg\bullet(\text{``}A_1\text{''})}\,\mathrm{Rup^n} \quad \ldots \quad \dfrac{\dfrac{\Pi_n}{\neg A_n}}{\neg\bullet(\text{``}A_n\text{''})}\,\mathrm{Rup^n}}{\dfrac{\Pi_M}{\bullet(\text{``}A\text{''})}} \qquad \dfrac{\bullet(\text{``}A\text{''})}{A}\,\mathrm{Rdw}$$

where each Π_i is a deduction of $\neg A_i$ from the set of assumptions Γ_O, Γ_{iM}; and Π_M is a deduction of $\bullet(\text{``}A\text{''})$ from $\Gamma_M, \neg\bullet(\text{``}A_1\text{''})\ldots\neg\bullet(\text{``}A_n\text{''})$ in M which discharges all the assumptions in Γ_{iM} not in Γ_M.

From the fact that $\bullet(\text{``}A\text{''})$ is derivable from $\Gamma_M, \neg\bullet(\text{``}A_1\text{''}), \ldots, \neg\bullet(\text{``}A_n\text{''})$ in M, we can infer (by propositional reasoning) that:

$$\Gamma_M \vdash_M \bullet(\text{``}A\text{''}) \vee \phi_1 \vee \ldots \vee \phi_n \tag{25}$$

where each ϕ_i is either the conjunction of the wffs of Γ_{iM} or $\bullet(\text{``}A_i\text{''})$. From (25) we have that for each wff $\psi_i \in (\Gamma_{iM})_O^{\vee} \cup \{A_i\}$

$$(\Gamma_M)_O^{\vee} \vdash_o A \vee \psi_1 \vee \ldots \vee \psi_n \tag{26}$$

By induction hypothesis we have that for each $1 \le i \le n$

$$\Gamma_O, (\Gamma_{iM})_O^{\vee} \vdash_o \neg A_i \tag{27}$$

Combining (26) and (27), by propositional reasoning we conclude that

$$\Gamma_O, (\Gamma_M)_O^{\vee} \vdash_o A.$$

(22) is a special case of (24), when $\Gamma_O = \Gamma_M = \emptyset$. In this case indeed we have that:

$$\vdash_{\mathrm{OM}} A \quad \Longrightarrow \quad (\emptyset)_O^{\vee} \vdash_o A$$

where $(\emptyset)_O^{\vee} = \{A_1 \vee \ldots \vee A_n : \bullet(\text{``}A_1\text{''}) \vee \ldots \vee \bullet(\text{``}A_n\text{''}) \in \mathrm{TH_{OM}}(M)\}$. Therefore, we have that $\mathrm{TH_{OM}}(O) \subseteq \mathrm{TH}(O + \{A_1 \vee \ldots \vee A_n : \bullet(\text{``}A_1\text{''}) \vee \ldots \vee \bullet(\text{``}A_n\text{''}) \in \mathrm{TH_{OM}}(M)\})$.

Viceversa. To prove that $\mathrm{TH}(O + \{A_1 \vee \ldots \vee A_n : \bullet(\text{``}A_1\text{''}) \vee \ldots \vee \bullet(\text{``}A_n\text{''}) \in \mathrm{TH_{OM}}(M)\}) \subseteq \mathrm{TH_{OM}}(O)$, it is enough to show that if $\bullet(\text{``}A_1\text{''}) \vee \ldots \vee \bullet(\text{``}A_n\text{''})$ is a theorem of the metatheory of OM than $A_1 \vee \ldots \vee A_n$ can be proved in the object theory of OM. The following is a proof:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\neg(A_1\vee\ldots\vee A_n)}{\neg A_n}}{\neg\bullet(``A_n")}\,\mathrm{Rup^n} \qquad \begin{array}{c}\dfrac{\dfrac{\dfrac{\neg(A_1\vee\ldots\vee A_n)}{\neg A_1}}{\neg\bullet(``A_1")}\,\mathrm{Rup^n}\quad \begin{array}{c}\vdots\\ \bullet(``A_1")\vee\bullet(``A_2")\vee\ldots\vee\bullet(``A_n")\end{array}}{\bullet(``A_2")\vee\ldots\vee\bullet(``A_n")}\\ \vdots\\ \bullet(``A_n")\end{array}}{\bot}\,\supset\mathrm{E}}{\bullet(``\bot")}\,\bot}{\bot}\,\mathrm{Rdw}}{A_1\vee\ldots\vee A_n}\,\bot
$$

This concludes the proof of (22).

Finally, (23) is a consequence of (21), and of the fact that $\mathrm{Rup^n} + \mathrm{Rdw} =_M \mathrm{Rup^n_r} + \mathrm{Rdw}$ (line 7 and column 13 of Table 1).

4.2 Fixpoint characterization

The methodology presented in Section 4.1 is rather ad hoc. There are no general criteria which can guide the choice of suitable axiom schemata. Indeed there are no special reasons, apart from their common use, for choosing (K), (nTbot), (Comp), and others. This method, furthermore, is not exhaustive, as there are combinations of reflection rules, such as, for instance, $\mathrm{Rup_r} + \mathrm{Rdw_r}$, which are have not been characterized. Our goal, in this subsection, is to provide a general methodology for characterizing $\mathrm{TH_{OM}}(O)$ and $\mathrm{TH_{OM}}(M)$. The idea is to consider a set of reflection rules as an operator on pairs of object theory and meta theory. The operator associated to the set of reflection rules (RR) extends the object theory by applying the reflection down rules of (RR) to the theorems of the metatheory, and the metatheory by applying the reflection up rules of (RR) to the theorems of the object theory. The set of object theorems and the set of the meta theorems of an OM pair composed of O and M connected by (RR) can be expressed, therefore, as the minimal fixpoint of the operator associated to (RR) that contains $\mathrm{TH}(O)$ and $\mathrm{TH}(M)$.

We start by considering the combination $\mathrm{Rup_r} + \mathrm{Rdw_r}$. We successively generalize this method to any set of reflection rules.

Given a reflection rule ρ, and a set of wffs $\Gamma \subseteq L_O \cup L_M$, $\rho(\Gamma)$ is the set of consequences of the application of ρ to the formulas of Γ. For instance if Γ is $\{\bullet(``p"), \neg\bullet(``q"), \bullet(``r")\wedge\bullet(``s"), p, \neg q\}$ $\mathrm{Rdw_r}(\Gamma) = \{p\}$, $\mathrm{Rdw^n_r}(\Gamma) = \{\neg q\}$ and $\mathrm{Rup}(\Gamma) = \{\bullet(``p"), \bullet(``\neg q")\}$. For a set of reflection rules (RR), we define $(\mathrm{RR})(\Gamma) = \bigcup_{\rho\in(\mathrm{RR})}\rho(\Gamma)$.

Theorem 15 Let OM be an OM pair composed of O and M connected by the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw_r}$. Then $\mathrm{TH_{OM}}(O)$ and $\mathrm{TH_{OM}}(M)$ are the smallest sets of wffs which satisfy the following equations:

$$x_O = \mathrm{TH}(O + \mathrm{Rdw_r}(x_M)) \tag{28}$$

$$x_M = \mathrm{TH}(M + \mathrm{Rup_r}(x_O)) \tag{29}$$

Proof Let us define the operatore $\mathfrak{O}_{\text{Rup}_r+\text{Rdw}_r}$ on a pair $\langle x_O, x_M\rangle \in 2^{L_O} \times 2^{L_M}$ as follows: $\mathfrak{O}_{\text{Rup}_r+\text{Rdw}_r}\langle x_O, x_M\rangle = \langle x'_O, x'_M\rangle$ where:

$$x'_O = \text{TH}(O + \text{Rdw}_r(x_M)) \tag{30}$$
$$x'_M = \text{TH}(M + \text{Rup}_r(x_O)) \tag{31}$$

$\mathfrak{O}_{\text{Rup}_r+\text{Rdw}_r}$ is monotonic. This implies that

$$\mathfrak{O}^{\omega}_{\text{Rup}_r+\text{Rdw}_r}\langle\emptyset,\emptyset\rangle = \bigcup_{k\geq 0} \mathfrak{O}^{k}_{\text{Rup}_r+\text{Rdw}_r}\langle\emptyset,\emptyset\rangle$$

is the least fixed point of $\mathfrak{O}_{\text{Rup}_r+\text{Rdw}_r}$. Let the *reflection degree* of a proof Π be the maximum number of reflection rules that occour in a thread of Π. Notice that $\mathfrak{O}^{k}_{\text{Rup}_r+\text{Rdw}_r}\langle\emptyset,\emptyset\rangle$ contains the object theorems and the meta theorems of OM that have a proof with reflection degree less then k. Since for any theorem of OM there is such a k, then

$$\langle \text{TH}_{\text{OM}}(O), \text{TH}_{\text{OM}}(M)\rangle = \mathfrak{O}^{\omega}_{\text{Rup}_r+\text{Rdw}_r}\langle\emptyset,\emptyset\rangle$$

This characterization can be trivially generalized to all the combinations of restricted reflection rules. For any set of reflection rules (RR), (RRup) is the set of reflection up rules in (RR), similarly, (RRdw) is the set of reflection down rules in (RR).

Theorem 16 Let OM be an OM pair composed of O and M connected by a set of restricted reflection rules (RR). $\text{TH}_{\text{OM}}(O)$ and $\text{TH}_{\text{OM}}(M)$ are the smallest sets of wffs which satisfy the following fixpoint equations:

$$x_O = \text{TH}(O + (\text{RRdw})(x_M)) \tag{32}$$
$$x_M = \text{TH}(M + (\text{RRup})(x_O)) \tag{33}$$

Restricted reflection rules confine the effect of the assumptions in a single theory. This means, for instance, that the assumptions done in the object theory do not add any theorem in the meta theory. Formally this corresponds to properties (34) and (35).

$$\Gamma_O, \Gamma_M \vdash_{\text{OM}} A_O \quad \text{iff} \quad \Gamma_O \vdash_{\text{OM}} A_O \tag{34}$$
$$\Gamma_O, \Gamma_M \vdash_{\text{OM}} A_M \quad \text{iff} \quad \Gamma_M \vdash_{\text{OM}} A_M \tag{35}$$

The former property holds for OM pairs with restricted reflection down rules, while the latter holds for OM pairs with restricted reflection up rules.

Unrestricted reflection rules, instead, make it possible for assumptions in a theory to affect provability in the other theory. For instance, the assumption $\bullet(\text{“}p\text{”})$ in the meta theory, forces p, and all its logical consequences, to be theorems in the object theory. Unrestricted reflection rules enable *hypothetical reasoning*, either on the object theory, or on the metatheory. An example of hypothetical reasoning on the object theory is the following: assume $\bullet(\text{“}p\text{”})$ in the metatheory, by unrestricted reflection down, infer p, then, by reasoning in the object theory, infer $p \vee p$, and by reflection up, conclude $\bullet(\text{“}p\vee p\text{”})$ in the meta theory. By means of this reasoning we have proved the implication $\bullet(\text{“}p\text{”}) \supset \bullet(\text{“}p \vee p\text{”})$. The characterization of the unrestricted reflection rules is based on the intuition that unrestricted reflection rules can be simulated by restricted reflection rules plus the implication theorems provable by hypothetical reasoning. This intuition allows us to generalize Theorem 16 to any set of reflection rules.

Let us introduce some new notation. For any set of reflection rule (RR), $(\mathrm{RRup_u})$ denotes the set of unrestricted reflection up of (RR), while $(\mathrm{RRdw_u})$ denotes the set of unrestricted reflection down of (RR). For any OM pair let us define H_O and H_M, as the sets of theorems provable in O and in M respectively, by hypothetical reasoning.

$$
\begin{aligned}
H_O &= \left\{ \bigwedge_{i=1}^{n} A_i \supset B \mid B \in (\mathrm{RRdw})(\mathrm{TH}(M + (\mathrm{RRup_u})(A_1, \ldots, A_n))) \right\} \\
H_M &= \left\{ \bigwedge_{i=1}^{n} A_i \supset B \mid B \in (\mathrm{RRup})(\mathrm{TH}(O + (\mathrm{RRdw_u})(A_1, \ldots, A_n))) \right\}
\end{aligned}
$$

Theorem 17 Let OM be an OM pair composed of O and M connected by a set of reflection rules (RR). $\mathrm{TH_{OM}}(O)$ and $\mathrm{TH_{OM}}(M)$ are the smallest sets of wffs which satisfy the following fixpoint equations:

$$
\begin{aligned}
x_O &= \mathrm{TH}(O + (\mathrm{RRdw})(x_M) + H_O) && (36) \\
x_M &= \mathrm{TH}(M + (\mathrm{RRup})(x_O) + H_M) && (37)
\end{aligned}
$$

If $(\mathrm{RRdw_u})$ is empty, then H_M is propositionally equivalent to $(\mathrm{RRup})(\mathrm{TH}(O))$, and if $(\mathrm{RRup_u})$ is empty, then H_O is propositionally equivalent to $(\mathrm{RRdw})(\mathrm{TH}(M))$. With restricted reflection rules, therefore, Theorem 17 reduces to Theorem 16.

Proof To prove Theorem 17 it is enough to show that any occurrence of an unrestricted reflection rule in a proof can be replaced by some form of hypothetical reasoning: we show the case of Rup; all the other cases are analogous. Let Π be a proof containing an application of Rup. Suppose that the premise of Rup depends on some set of assumptions in the object theory. Being Π a proof, we have that there must be an application of a reflection down occurring below that of the Rup, which is necessary to discharge the assumptions. This implies that Π is of the following form:

$$
\begin{array}{c}
\dfrac{\begin{array}{c}\Pi_1 \\ A_1\end{array}}{\bullet(\text{``}A_1\text{''})}\,\mathrm{Rup} \quad \cdots \quad \dfrac{\begin{array}{c}\Pi_n \\ A_n\end{array}}{\bullet(\text{``}A_n\text{''})}\,\mathrm{Rup} \\
\Pi_M \\
\dfrac{C_M}{B_O}\ \rho \in (\mathrm{RRdw}) \\
\Pi'
\end{array}
$$

where Π_M is a proof of C_M in $M + \{\bullet(\text{``}A_1\text{''}), \ldots, \bullet(\text{``}A_n\text{''})\}$. The application of Rup can be replaced with the axiom $\bigwedge_{i=1}^{n} A_i \supset B_O \in H_O$ as follows:

$$
\begin{array}{c}
\dfrac{\bigwedge_{i=1}^{n} A_i \supset B_O \quad A_1 \quad \ldots \quad A_n}{B_O} \\
\Pi'
\end{array}
$$

Theorem 17 provides the answer to the problem of characterizing the object theory and the metatheory generated by *any* set of reflection rules.

4.3 Schematic vs. fixpoint characterizations

Beside being uniform and general, the fixpoint methodology allows us to characterizes the behavior of a certain set of reflection rules which are not characterizable via schematic characterization. To state this formally, we have to better qualify the notion of *schematic characterization.* Let $P_1, P_2, \ldots$ be an infinite set of propositional parameters contained in L_0. Each P_i is not a simple propositional letter, but it must be intended as a propositional variable ranging over L_O formulas. A *schematic formula* is a formula in the language of the metatheory possibly extended with the set of parameters. Examples of schematic formulas are the axioms (K), (Cons) and (nTbot). $\Phi[P_1, \ldots, P_n]$ denotes a schematic formula that contains the parameters $P_1, \ldots, P_n$. An *instance* $\Phi[A_1, \ldots, A_n]$, of the schematic formula $\Phi[P_1, \ldots, P_n]$, is an L_M formula obtained by replacing each parameter P_i with the object formula A_i.

Definition 18 (Schematic characterization) A set (RR) of reflection rules has a *schematic characterization* if there is a schematic formula $\Phi[P_1, \ldots, P_n]$ such that, for any OM pair $\mathrm{OM} = \langle O, M, (\mathrm{RR})\rangle$,

$$\mathrm{TH}_{\mathrm{OM}}(M) = \mathrm{TH}(M + (\mathrm{RR})(\mathrm{TH}(O)) + \{\Phi[A_1, \ldots, A_n] : A_i \in L_O\}) \qquad (38)$$

The literature contains many approaches where a metatheory is defined in terms of schematic formulas. For instance, in metalogic programming [22], metainterpreters are very similar to the schematic formula (K); in provability logic [32], the metatheory is built by the axiom schema $\Box(\Box A \supset A) \supset \Box A)$, and the necessitation rule $\frac{\vdash A}{\vdash \Box A}$; Feferman's local reflection principles [12] is also a schematic formula. Now the interesting question is whether reflection rules allow us to generate meaningful and interesting metatheories which are not definable by means of schematic formulas. Let's first compare schematic and fixpoint characterization.

Lemma 19 If a set of reflection rules (RR) has an axiomatic characterization, then there is an $i \geq 0$, such that, for all OM pairs composed of O and M connected by (RR), O_i and M_i are the solution of equations (37) and (36), where $O_0 = \mathrm{TH}(O)$, $M_0 = \mathrm{TH}(M)$ and

$$\begin{aligned} M_{i+1} &= \mathrm{TH}(M + (\mathrm{RRup})(O_i)) \\ O_{i+1} &= \mathrm{TH}(O + (\mathrm{RRdw})(M_i)) \end{aligned}$$

Proof (Lemma 19) Let $\underline{O}$ be propositional theory with the empty set of axioms on a language L, and let $\underline{M}$ be a propositional theory with the empty set of axioms, on the propositional meta language of L. Let $\underline{\mathrm{OM}} = \langle \underline{O}, \underline{M}, (\mathrm{RR})\rangle$. Let $\Phi[P_1, \ldots, P_n]$ be a schema for the set of reflection rules (RR). Let Π be a proof of $\Phi[P_1, \ldots, P_n]$ in $\underline{\mathrm{OM}}$. Being finite, Π contains a finite number of applications of reflection rules, and therefore there is a natural number i such that $\Phi[P_1, \ldots, P_n] \in \underline{M}_i$ Since the inference rules of the object and the meta theories and the reflection rules (RR) are schematic on the object language[3], the proof of any instance $\Phi[A_1, \ldots, A_n]$ can be obtained by replacing $P_1 \ldots, P_n$ with $A_1, \ldots, A_n$ in Π. This implies that $\underline{M}_i$ contains also all the instances of $\Phi[P_1, \ldots, P_n]$. By monotonicity of the derivability relation, we have that

[3]An inference rule ρ is schematic on a language L , if the result of substituting L formulas for parameters in an application of ρ, is still an application of ρ

$\Phi[A_1, \ldots, A_n]$ belongs to M_i, for any metatheory M. By the definition of schematic representation, (equation (38)), we have that $\mathrm{TH}_{\mathrm{OM}}(M) \subseteq M_i \subseteq \mathrm{TH}_{\mathrm{OM}}(M)$ which implies $\mathrm{TH}_{\mathrm{OM}}(M) = M_i$

Theorem 20 The sets of reflection rules $\mathrm{Rup_r} + \mathrm{Rdw_r}$, and $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw_r^n}$ do not have a schematic characterization.

Proof (Theorem 20) Let $\underline{\mathrm{OM}}$, be the OM pair defined in the proof of Lemma 19, with (RR) = $\mathrm{Rup_r} + \mathrm{Rdw_r}$. Suppose by contradiction that there is a schematic characterization of $\mathrm{Rup_r} + \mathrm{Rdw_r}$. By Lemma 19 let i be the integer such that O_i and M_i are the solution of equations (37) and (36). Notice that such an $i = 1$ as in $\underline{OM}$, $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M) = O + \mathrm{Rup}(\mathrm{TH}(O))$. On the other hand it is easy to find an OM pair where M_1 is not a fixpoint. Consider, for instance, the OM pair obtained by extending $\underline{OM}$ by the meta axiom $\bullet(\text{“}p\text{”})$. We have that $\bullet(\text{“}p \wedge p\text{”}) \in M_2$, but $\bullet(\text{“}p \wedge p\text{”}) \notin M_1$. Indeed to deduce $\bullet(\text{“}p \wedge p\text{”})$ from $\bullet(\text{“}p\text{”})$ we need two applications of bridge rules, first $\mathrm{Rdw_r}$ and then $\mathrm{Rup_r}$.) This contradicts $\mathrm{TH}(M_1) = \mathrm{TH}_{\mathrm{OM}}(M)$. The case of $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw_r^n}$ can be proved in the same way.

From the proof of Theorem 20, we conclude that there does not exist an axiom schema which characterizes the meta theories generated by $\mathrm{Rup_r} + \mathrm{Rdw_r}$. Notice that the meta theory generated by $\mathrm{Rup_r} + \mathrm{Rdw_r}$ is very important. First because it is the most used and quoted. Second, because it is the minimal meta theory which is sound and complete for provability in the object theory, that is, it is the smallest M such that it is in the relation "$\vdash_M$ *theorem*(“A”) iff $\vdash_O A$" with O. The metatheory generated by $\mathrm{Rup_r} + \mathrm{Rdw_r} + \mathrm{Rdw_r^n}$ is interesting because it is the smallest metatheory for provability extended with a *negation as failure* rule such as: *"if in the metatheory it is provable that A is not a theorem, then ¬A is a theorem"*.

Table 2 summarizes the results presented in this section.

5 Duality

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 [23] and [11] respectively. The results provided in the previous sections mainly concern provability and derivability in OM pairs. In order to extend these results we define a duality principle that preserves derivability. According to this principle:

> *if a formula is derivable in an OM pair from a set of assumptions, then the dual formula is derivable in the dual OM pair from the dual set of assumptions*

Definition 21 (Dual reflection rule) For any positive [negative] reflection rule ρ, *dual*(ρ) is the corresponding negative [positive] reflection rule.

Definition 22 (Dual formula) For any formula $A \in L_M$, *dual*(A) is defined accordingly to 1–4:

(RR)	Meta schema	Object schema
$\emptyset$	$\emptyset$	$\emptyset$
$\mathrm{Rdw_r}$	$\emptyset$	Rdw(TH(M))
$\mathrm{Rup_r}$	Rup(TH(O))	$\emptyset$
Rdw	$\emptyset$	Rdw(TH(M))
Rup	Rup(TH(O))	$\emptyset$
$\mathrm{Rup_r Rdw_r}$	Fixpoint	Fixpoint
$\mathrm{RdwRup_r^n}$	$\mathrm{Rup^n}$(TH(O)) + (Cons)	Rdw($\mathrm{TH_{OM}}$(M))
$\mathrm{RdwRup^n}$	$\mathrm{Rup^n}$(TH(O)) + (Cons)	$\bigvee A_i : \bigvee \bullet(``A_i") \in \mathrm{TH_{OM}}(M)\}$
$\mathrm{Rup_r Rdw}$	Rup(TH(O)) + (K)	Rdw($\mathrm{TH_{OM}}$(M))
$\mathrm{Rup_r RdwRup_r^n}$	Rup(TH(O)) + (K) + (nTbot)	Rdw($\mathrm{TH_{OM}}$(M))
$\mathrm{Rup_r RdwRdw_r^n}$	Fixpoint	Fixpoint
$\mathrm{Rup_r RdwRup^n}$	Rup(TH(O)) + (K) + (nTbot)	Rdw($\mathrm{TH_{OM}}$(M))
$\mathrm{Rup_r RdwRdw^n}$	Rup(TH(O)) + (K) + (Comp)	Rdw($\mathrm{TH_{OM}}$(M))
RupRdw	Rup(TH(O)) + (K) + (Comp)	Rdw($\mathrm{TH_{OM}}$(M))
$\mathrm{RupRdwRup_r^n}$	Rup(TH(O)) + (K) + (nTbot) + (Comp)	Rdw($\mathrm{TH_{OM}}$(M))

Table 2: Schematic characterization of the theories generated by sets of reflection rules

1. $dual(A) = A$, A does not contain any occurrence of $\bullet$;
2. $dual(\bullet(``A")) = \neg \bullet (``\neg A")$;
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 meta properties. To define $dual(\bullet(``A"))$ we reason as follows: since $\bullet(``A")$ is derivable by Rup from A, $dual(\bullet(``A"))$ must be derivable by *dual*(Rup) (i.e. $\mathrm{Rup^n}$) from $A$. This implies that $dual(\bullet(``A"))$ must be an element of the set of formulas derivable from A by $\mathrm{Rup^n}$. In symbols $dual(\bullet(``A")) \in \{\neg\bullet(``B") : A \vdash_{\mathrm{O}} \neg B\}$. We have chosen $dual(\bullet(``A")) = \neg \bullet (``\neg A")$.

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 $\bullet(``B")$, $\bullet(``B")$ must be equivalent to $\neg\neg \bullet (``\neg\neg B")$, 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:

$$\vdash_{\mathrm{OM}} \bullet(``A") \equiv \bullet(``\neg\neg A") \quad (39)$$

We call the OM pairs which satisfy Condition (39), *classical*.

Definition 23 (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 $dual(\Omega_M)$ and by adding $\bullet(``A") \equiv \bullet(``\neg\neg A")$.

In the definition of dual OM pair, we force $dual(OM)$ to be classical by explicitly adding the axiom schema $\bullet(\text{“}A\text{”}) \equiv \bullet(\text{“}\neg\neg A\text{”})$. In many cases this is not necessary, e.g., for the OM pairs with reflection rules stronger or equal to $\mathrm{Rup_r} + \mathrm{Rdw}$. There are cases, however, where OM is classical but $dual(OM)$ is not. This should be result clearer by observing that the dual of $\bullet(\text{“}A\text{”}) \equiv \bullet(\text{“}\neg\neg A\text{”})$ is

$$\neg \bullet (\text{“}\neg A\text{”}) \equiv \neg \bullet (\text{“}\neg\neg\neg A\text{”})$$

which is equivalent to $\bullet(\text{“}\neg A\text{”}) \equiv \bullet(\text{“}\neg\neg\neg A\text{”})$, but which is not equivalent to $\bullet(\text{“}A\text{”}) \equiv \bullet(\text{“}\neg\neg A\text{”})$.

Lemma 24 Let OM be a classical OM pair, Then:

1. $\vdash_{\mathrm{OM}} A \equiv dual(dual(A))$;
2. $\mathrm{OM} =_D dual(dual(\mathrm{OM}))$.

Proof $\vdash_{\mathrm{OM}} A \equiv dual(dual(A))$ is provable by induction on the complexity of A. If A is $\bullet(\text{“}B\text{”})$, then $dual(dual(A))$ is $\neg\neg \bullet (\text{“}\neg\neg B\text{”})$. The fact that OM is classical implies that $\vdash_{\mathrm{OM}} \bullet(\text{“}B\text{”}) \equiv \neg\neg \bullet (\text{“}\neg\neg B\text{”})$, namely, that A is equivalent in OM to $dual(dual(A))$. The step cases are straightforward as $dual(.)$ distributes over connectives. (2) is a direct consequence of (1).

Theorem 25 (Duality preserves derivability) Let OM be a classical OM pair, Γ_O, A_O and Γ_M, A_M be a set of object wffs and a set of meta wffs respectively. Then

$$\Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_O \iff \Gamma_O, dual(\Gamma_M) \vdash_{dual(\mathrm{OM})} A_O \tag{40}$$

$$\Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_M \iff \Gamma_O, dual(\Gamma_M) \vdash_{dual(\mathrm{OM})} dual(A_M) \tag{41}$$

Proof For the proof of the left-to-right direction we proceed by induction on the complexity of the deductions in OM.

Base case If Π is the assumption A_O, then it belongs to Γ_O; if Π is the axiom A_O, then A_O is also an axiom of $dual(\mathrm{OM})$. In both cases the right side of (40) holds. If Π is the assumption A_M, then it belongs to Γ_M. Since $dual(A_M)$ belongs to $dual(\Gamma_M)$ the right side of (41) holds. If Π is the axiom A_M, then $dual(A_M)$ is an axiom of $dual(\mathrm{OM})$. This implies the right side of (41).

Step Case Let us consider the case of $\supset$I in the meta theory. The cases of introduction and elimination of the other connectives both in the object theory and in the meta theory are similar. If Π ends with an application of $\supset$I in the meta theory, then it is of the form:

$$\frac{\begin{array}{c}[B]\\ \Pi' \\ C\end{array}}{B \supset C} \supset\mathrm{I}_M$$

By applying the induction hypothesis to Π' we obtain that:

$$\Gamma_O, dual(\Gamma_M), dual(B) \vdash_{\mathrm{OM}} dual(C)$$

(RR)	Meta schema	Object schema
$\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\mathrm{Rdw}^{\mathrm{n}}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}(\mathrm{K})$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$
$\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\mathrm{Rdw}^{\mathrm{n}}\mathrm{Rup}_{\mathrm{r}}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}((\mathrm{K}) + (\mathrm{nTbot}))$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$
$\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\mathrm{Rdw}^{\mathrm{n}}\mathrm{Rdw}_{\mathrm{r}}$	Fixpoint	Fixpoint
$\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\mathrm{Rdw}^{\mathrm{n}}\mathrm{Rup}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}((\mathrm{K}) + (\mathrm{nTbot}))$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$
$\mathrm{Rup}^{\mathrm{n}}_{\mathrm{r}}\mathrm{Rdw}^{\mathrm{n}}\mathrm{Rdw}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}((\mathrm{K}) + (\mathrm{Comp}))$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$
$\mathrm{Rup}^{\mathrm{n}}\mathrm{Rdw}^{\mathrm{n}}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}((\mathrm{K}) + (\mathrm{Comp}))$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$
$\mathrm{Rup}^{\mathrm{n}}\mathrm{Rdw}^{\mathrm{n}}\mathrm{Rup}_{\mathrm{r}}$	$\mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}((\mathrm{K}) + (\mathrm{nTbot}) + (\mathrm{Comp}))$	$\mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))$

Table 3: Schematic characterization of the theories generated by sets of dual reflection rules

and therefore $\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathrm{OM}} \mathit{dual}(B) \supset \mathit{dual}(C)$. By duality, $\mathit{dual}(B) \supset \mathit{dual}(C)$ is equal to $\mathit{dual}(B \supset C)$ and therefore (41) holds.

If Π ends with an application of a reflection rule, then it is of one of the following forms:

$$\begin{array}{cccc}
\dfrac{\begin{array}{c}\Gamma_O, \Gamma_M \\ \Pi \\ A\end{array}}{\bullet(\text{“}A\text{”})}\ \mathrm{Rup} &
\dfrac{\begin{array}{c}\Gamma_O, \Gamma_M \\ \Pi \\ \neg A\end{array}}{\neg\bullet(\text{“}A\text{”})}\ \mathrm{Rup}^{\mathrm{n}} &
\dfrac{\begin{array}{c}\Gamma_O, \Gamma_M \\ \Pi \\ \bullet(\text{“}A\text{”})\end{array}}{A}\ \mathrm{Rdw} &
\dfrac{\begin{array}{c}\Gamma_O, \Gamma_M \\ \Pi \\ \neg\bullet(\text{“}A\text{”})\end{array}}{\neg A}\ \mathrm{Rdw}^{\mathrm{n}} \\
(a) & (b) & (c) & (d)
\end{array}$$

(The cases of restricted reflection rules are special cases in which Γ_O or Γ_M are empty). We consider the case of Rup, the other cases are similar. If Π is of the form (a), by induction, we have that A and therefore $\neg\neg A$ is derivable from $\Gamma_O, \mathit{dual}(\Gamma_M)$. By $\mathrm{Rup}^{\mathrm{n}}$, which is an inference rule of $\mathit{dual}(\mathrm{OM})$, we have that $\neg\bullet(\text{“}\neg A\text{”})$ is derivable in $\mathit{dual}(\mathrm{OM})$ from $\Gamma_O, \mathit{dual}(\Gamma_M)$. Therefore, we have that $\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathit{dual}(\mathrm{OM})} \neg\bullet(\text{“}\neg A\text{”})$.

To prove the right-to-left direction of Theorem 25, we apply the left-to-right direction of the same theorem to $\mathit{dual}(\mathrm{OM})$ obtaining the following:

$$\begin{array}{rcl}
\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathit{dual}(\mathrm{OM})} A_O & \Longrightarrow & \Gamma_O, \mathit{dual}(\mathit{dual}(\Gamma_M)) \vdash_{\mathit{dual}(\mathit{dual}(\mathrm{OM}))} A_O \\
\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathit{dual}(\mathrm{OM})} A_M & \Longrightarrow & \Gamma_O, \mathit{dual}(\mathit{dual}(\Gamma_M)) \vdash_{\mathit{dual}(\mathit{dual}(\mathrm{OM}))} \mathit{dual}(\mathit{dual}(A_M))
\end{array}$$

By Lemma 24 we have that $\mathit{dual}(\mathit{dual}(OM))$ is equivalent to OM. Since OM is classical, $\mathit{dual}(\mathit{dual}(\Gamma_M))$ is equivalent in OM to Γ_M, $\mathit{dual}(\mathit{dual}(A_M))$ is equivalent to A_M. We can therefore conclude that:

$$\begin{array}{rcl}
\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathit{dual}(\mathrm{OM})} A_O & \Longrightarrow & \Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_O \\
\Gamma_O, \mathit{dual}(\Gamma_M) \vdash_{\mathit{dual}(\mathrm{OM})} \mathit{dual}(A_M) & \Longrightarrow & \Gamma_O, \Gamma_M \vdash_{\mathrm{OM}} A_M
\end{array}$$

Theorem 25 allows us to extend the schematic characterization provided in Subsection 4.1 to the dual combination of reflection rules.

Theorem 26 (Dual of Theorem 13) Let OM be an OM pair composed of O and M connected by the set of reflection rules (RR). then:

$$\begin{array}{rcl}
\mathrm{TH}_{\mathrm{OM}}(O) & = & \mathrm{TH}(O + \mathrm{Rdw}^{\mathrm{n}}(\mathrm{TH}_{\mathrm{OM}}(M))) \\
\mathrm{TH}_{\mathrm{OM}}(M) & = & \mathrm{TH}(M + \mathrm{Rup}^{\mathrm{n}}(\mathrm{TH}(O)) + \mathit{dual}(\Gamma))
\end{array}$$

where:

1. If (RR) is $\mathrm{Rup^n_r} + \mathrm{Rdw^n}$, then $\Gamma = (\mathrm{K})$;
2. If (RR) is $\mathrm{Rup^n_r} + \mathrm{Rdw^n} + \mathrm{Rup_r}$, then $\Gamma = (\mathrm{K}) \cup (\mathrm{nTbot})$;
3. If (RR) is $\mathrm{Rup^n} + \mathrm{Rdw^n}$, then $\Gamma = (\mathrm{K}) \cup (\mathrm{Comp})$;
4. If (RR) is $\mathrm{Rup^n} + \mathrm{Rdw^n} + \mathrm{Rup_r}$ then $\Gamma = (\mathrm{K}) \cup (\mathrm{Comp}) \cup (\mathrm{nTbot})$.

Proof To prove Theorem 26 it is sufficient to prove that if (RR) $\geq_D \mathrm{Rup^n_r} + \mathrm{Rdw^n}$, then (RR) generates a classical OM pair. In other words, we prove that the combination $\mathrm{Rup^n_r} + \mathrm{Rdw^n}$ generates a classical OM pair. A proof of $\bullet(``A") \equiv \bullet(``\neg\neg A")$ in an OM pair generated by $\mathrm{Rup^n_r} + \mathrm{Rdw^n}$ is the following:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\neg\bullet(``\neg\neg A")}{\neg\neg\neg A}\,\mathrm{Rdw^n}}{\neg A}}{\neg\bullet(``A")}\,\mathrm{Rup^n_r}\quad \bullet(``A")}{\bot}\,\supset\mathrm{E}}{\bullet(``\neg\neg A")}\,\bot}{\bullet(``A")\supset\bullet(``\neg\neg A")}\,\supset\mathrm{I} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\neg\bullet(``A")}{\neg A}\,\mathrm{Rdw^n}}{\neg\neg\neg A}}{\neg\bullet(``\neg\neg A")}\,\mathrm{Rup^n_r}\quad \bullet(``\neg\neg A")}{\bot}\,\supset\mathrm{E}}{\bullet(``A")}\,\bot}{\bullet(``\neg\neg A")\supset\bullet(``A")}\,\supset\mathrm{I}}{\bullet(``A")\equiv\bullet(``\neg\neg A")}\,\wedge\mathrm{I}$$

We can therefore expand Table 2, by adding the proof theoretic characterization of the theories generated by dual reflection rules shown in Table 3.

An intuitive interpretation of *dual*$((K))$, *dual*(Comp), and *dual*(nTbot) is possible in terms of consistency. Notice that consistency is the dual concept of provability, and that (K), (nTbot) and (Comp) can be interpreted in terms of provability. The axiom schema *dual*((K)) is equivalent to

$$\bullet(``A") \supset (\neg\bullet(``A \wedge B") \supset \bullet(``\neg B")) \qquad (\mathrm{K^n})$$

Reading $\bullet(``A")$ as "A is consistent with the object theory", the instantiation of $(\mathrm{K^n})$ with A to $\top$ states that "If the object theory is consistent and B is inconsistent with it, then $\neg B$ is consistent with the object theory". The axiom schema *dual*((nTbot)) is equivalent to:

$$\bullet(``\top") \qquad (\mathrm{nTbot}^n)$$

(nTbot^n) can be read as *"$\top$ is consistent with the object theory"*, which is equivalent to say that the object theory is consistent. Finally *dual*((Comp)) is equivalent to

$$\bullet(``A") \vee \bullet(``\neg A") \qquad (\mathrm{Comp^n})$$

which is equal to (Comp) and can be read as *"either A or $\neg A$ is consistent with the object theory"*.

6 Case studies

The results obtained in the previous sections can be specialized to various subclasses of metatheories. In particular we consider the class of OM pairs with an empty metatheory (i.e., a metatheory with no axioms), the class of OM pairs with a Horn metatheory (i.e., a metatheory whose axioms are Horn clauses), the class of OM pairs where the object and metatheories are amalgamated, and finally the class of OM pairs with first order theories.

6.1 Empty metatheory

In an OM pair with an empty metatheory (i.e. if $\Omega_M = \emptyset$), all the theorems in $\mathrm{TH}_{\mathrm{OM}}(M)$ which are not tautologies, are generated by applying a set of reflection rules to the theorems of the object theory. The metatheory is therefore *completely generated* by the object theory via reflection rules. This class of OM pairs corresponds to those approaches to meta reasoning where the metatheory is defined on the basis of some object theory. In these approaches, the resulting metatheory is proved to be correct and/or complete with respect to the original object theory. In OM pairs, reflection rules guarantee that the metatheory generated in the OM pair is correct and/or complete with respect to $\mathrm{TH}_{\mathrm{OM}}(O)$, but not with respect to the original object theory $\mathrm{TH}(O)$. To show the correspondence between the generative approach described above and OM pairs with empty metatheory, we have to prove that, when the metatheory is empty, $\mathrm{TH}_{\mathrm{OM}}(O) = \mathrm{TH}(O)$.

Theorem 27 Let OM be an OM pair composed of O and an empty metatheory M, connected by any set of reflection rules. Then for any set of object wffs Γ, Σ, A:

1. If $\bullet(\text{“}\Gamma\text{”}) \vdash_{\mathrm{OM}} \bigvee \bullet(\text{“}\Sigma\text{”})$, then $\Gamma \vdash_{\mathrm{OM}} \bigvee \Sigma$;
2. if $\Gamma \vdash_{\mathrm{OM}} A$, then $\Gamma \vdash_{\mathrm{O}} A$.

where, if Σ is the empty set, then $\bigvee \Sigma$ and $\bigvee \bullet(\text{“}\Sigma\text{”})$ are equal to $\bot$.

Proof For any deduction Π in OM, let $\overline{\Pi}^O$ be the formula tree obtained by replacing each occurrence of $\bullet(\text{“}A\text{”})$ in Π with A, and by removing the consequences of all the applications of the reflection rules. By induction on the structure of Π we can prove that, if Π is a deduction of A from Γ, than $\overline{\Pi}^O$ is a deduction of $\overline{A}^O$ from $\overline{\Gamma}^O$, where $\overline{A}^O$ and $\overline{\Gamma}^O$ are obtained by performing the same process as in $\overline{\Pi}^O$ (see [31] for a detailed proof of this fact). Let Π be a deduction of $\bot$ in OM from $\bullet(\text{“}\Gamma\text{”})$ and $\neg\bullet(\text{“}\Sigma\text{”})$. $\overline{\Pi}^O$ is a deduction of $\bot$ from Γ and $\neg\Sigma$. This proves (1). (2) can be proved in the same way. Let Π be a deduction of A from Γ in OM. Then $\overline{\Pi}^O$ is also a deduction of A from Γ. Since it does not contain any occurrence of formulas of the meta language then $\overline{\Pi}^O$ is a deduction in O. Therefore we can conclude that $\Gamma \vdash_{\mathrm{O}} A$.

A consequence of Theorem 27 is that, for any pair of combinations of reflection rules $(\mathrm{RR})_1$ and $(\mathrm{RR})_2$, we have that $(\mathrm{RR})_1 =_O (\mathrm{RR})_2 =_O \emptyset$. A second consequence is that restricted reflection down rules are redundant. More precisely $\mathrm{Rdw_r}$ and $\mathrm{Rdw_r^n}$ are admissible rules in any OM pair with an empty metatheory.

6.2 Horn metatheory

We say that M is a horn metatheory when its axioms are horn formulas of the following form:

$$p_1 \wedge \ldots p_m \supset p \tag{42}$$
$$p_1 \wedge \ldots \wedge p_m \wedge \bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”}) \supset \bullet(\text{“}A\text{”}) \tag{43}$$
$$p_1 \wedge \ldots \wedge p_m \wedge \bullet(\text{“}A_1\text{”}) \wedge \ldots \wedge \bullet(\text{“}A_n\text{”}) \supset \bot \tag{44}$$

with $m, n \geq 0$, and p an atomic formula which is not of the form $\bullet(\text{“}A\text{”})$. When the metapredicate "$\bullet$" represents provability, horn meta axioms have the following intuitive interpretation: (43), with $n > 0$, represents an inference rule of the object theory with the applicability condition $p_1 \wedge \ldots \wedge p_m$. With $n = 0$, (43) represents the fact that A is an axiom of the object theory under the condition $p_1 \wedge \ldots \wedge p_m$. (44) represents a consistency constraint on the object theory.

Theorem 28 Let OM be an OM pair composed of a horn metatheory M and an object theory O, connected either by $\mathrm{Rup_r} + \mathrm{Rdw_r}$, or by $\mathrm{Rup_r} + \mathrm{Rdw}$, or by $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$. Let OM' be the OM pair obtained by removing from the metatheory of OM the axioms of the form (43), and by adding, to the object theory, the corresponding inference rule:

$$\frac{A_1 \quad \ldots \quad A_n}{A}\ \rho_{(43)}$$

with the following application restriction: each A_i does not depend from any assumptions in O, and each p_i is provable in M. Then $\mathrm{OM} =_O \mathrm{OM}'$.

Proof Let consider the case of $\mathrm{Rup_r} + \mathrm{Rdw_r}$. We prove that $OM \leq_O \mathrm{OM}'$. Let $A \in \mathrm{TH_{OM}}(O)$. By Theorem 12, we have that there is an i such that, $A \in \mathrm{TH}(O_{i+1}) = \mathrm{TH}(O + \mathrm{Rdw}(\mathrm{TH}(M_i)))$. Suppose by induction that, for any $j < i$, $\mathrm{TH}(O_j)$ is a subset of the object theorems of OM'. Let us prove that $\mathrm{TH}(O_{i+1})$ is also a subset of the object theorems of OM'. If $A \in \mathrm{Rdw}(\mathrm{TH}(M_i))$, then $\bullet(\text{“}A\text{”}) \in \mathrm{TH}(M_i) = \mathrm{TH}(M + \mathrm{Rup}(\mathrm{TH}(O_{i-1})))$. This implies that there is a non empty set of axioms of the form (43) such that:

1. $p_1, \ldots, p_n$ are provable from M (notice that p_i can be proved without using formulas of the form $\bullet(\text{“}A\text{”})$, as p_i can be the consequence of a horn axiom in which $\bullet$ does not appear), and
2. each $\bullet(\text{“}A_i\text{”})$ of the left side of (43) is provable by a subset of such axioms.

By induction on the number of axioms, it is easy to show that each of $A_1, \ldots, A_n$ is provable in the object theory of OM'. The inference rule $\rho_{(43)}$ allows us, therefore, to infer that A belongs to the object theory of OM'.

To prove $\mathrm{OM}' \leq_O \mathrm{OM}$, it is sufficient to show that any application of the inference rule $\rho_{(43)}$, can be replaced by deduction in OM, that uses $\mathrm{Rup_r}$, the meta axiom (43) and $\mathrm{Rdw_r}$. The proof of the other cases are trivial as we can exploit the schematic characterization of M.

The previous theorem states that any object theory can be specified by a horn metatheory. For purpose of object theory specification, horn metatheories have been largely studied in the literature. See, for instance, [24] in the area of metalogic programming, and [5] in the area of metalogical framework, and the FOL metatheory as described in [36, 19]. Many of the ideas and results which have been developed in these approaches, can be reproduced in the OM pair framework. Consider for instance the meta predicate $demo_L$, as defined in [24] for some object level logic L. $demo_L$ is extensionally equivalent to the predicate $\bullet$ of the metatheory of the OM pair composed of L and an empty metatheory connected by $\text{Rup}_\text{r} + \text{Rdw}$. That is, given a program P and its meta program MP, any ground formula A is provable in MP if and only if the corresponding formula $A_\bullet^{demo}$ is provable in the metatheory of OM (where $A_\bullet^{demo}$ stands for the formula A in which all its atomic subformulas $demo(B)$ are substituted with $\bullet(\text{"}B\text{"})$).

6.3 Amalgamated object and metatheory theory

In OM pairs the object theory and the metatheory are kept distinct, but there is a lot of work in which the metatheory is either embedded or the same as the object theory. These situations are usually addressed as amalgamating object and metatheory. Here we show how amalgamated object and metatheory can be reconstructed in OM pairs.

Consider for instance the work in [8]. In this work the authors amalgamate an object language L_O, a consequence relation $\vdash_\text{O}$, a meta language L_M, and a consequence relation $\vdash_\text{M}$. $\vdash_\text{O}$ and $\vdash_\text{M}$ are connected by two inference rules which are a variant of Rup_r and Rdw_r. Following [8], we suppose that 1 $L_O = L_M$ and 2 $\vdash_\text{O}=\vdash_\text{M}$. These two conditions can be easily imposed in our framework. Requirement 1 states that L_O is the language of the metatheory. This implies that L_O contains its propositional meta language, i.e. $\bullet(\text{"}L_O\text{"}) \subseteq L_O$. Such a language can be obtained simply by adding the following formation rule:

1. if A is a wff then $\bullet(\text{"}A\text{"})$ is an atomic ground wff;

Requirement 2 states that the provability relation of the object theory is the same as the provability relation of the metatheory. Following our approach, we impose this constraint via rules which work between O and M. Intuitively we need a rule which exports theorems from the object theory to the metatheory and another which works in the opposite direction. Since object and metatheory have the same language, to keep things distinct, from now on we write $O : A$ or $M : A$ depending on whether $A \in L_O$ or $A \in L_M$. The rules which impose Requirement 2 are the following.

2.
$$\frac{O : A}{M : A}\ (\text{O}\subseteq\text{M})_r \qquad \frac{M : A}{O : A}\ (\text{M}\subseteq\text{O})_r$$

RESTRICTION: Both rules can be applied whenever their premises does not depend on assumptions in the same theory.

$(\text{O}\subseteq\text{M})_r$ forces $\vdash_\text{O}\subseteq\vdash_\text{M}$, while $(\text{M}\subseteq\text{O})_r$ forces $\vdash_\text{M}\subseteq\vdash_\text{O}$.

As far as we can see, many of the forms of amalgamation which can be found in the literature can be reproduced in the OM pair framework by using suitable combinations of reflection rules, $(\text{O}\subseteq\text{M})_r$ and $(\text{M}\subseteq\text{O})_r$. As an example, consider the rule of reflection down of the amalgamated system described in [1]:

$$\frac{\Gamma_1 \vdash PR(\text{``}\Gamma_2\text{''}, \text{``}A\text{''})}{\Gamma_1 \cup \Gamma_2 \vdash A} \text{ 3-reflect-down}$$

Let OM be an OM pair composed of an empty first order object theory and an empty first order metatheory, both theories using the same language L. Let the reflection rules of OM be Rup_r, Rdw, $(\text{O}\subseteq\text{M})_r$ and $(\text{M}\subseteq\text{O})_r$. Let us consider the following mapping $(.)^*$:

1. A^* is A if it does not contain any meta predicate PR;
2. $(.)^*$ distributes over connectives and quantifiers;
3. $(PR(\text{``}\Gamma\text{''}, \text{``}A\text{''}))^* = \bullet(\text{``} \wedge \Gamma^* \supset A^*\text{''})$.

The deduction in OM corresponding to 3-reflect-down can be defined as follows. Suppose $M : \Gamma_1^* \vdash M : \bullet(\text{``}\wedge\Gamma_2^* \supset A^*\text{''})$ and let $\Pi$ be a deduction of $M : \bullet(\text{``}\wedge\Gamma_2^* \supset A^*\text{''})$ from $M : \Gamma_1^*$. A deduction of $M : A^*$ from $M : \Gamma_1 \cup \Gamma_2$ is the following:

$$\dfrac{\dfrac{\dfrac{\begin{matrix} M : \Gamma_1^* \\ \Pi \\ M : \bullet(\text{``} \wedge \Gamma_2^* \supset A^*\text{''}) \end{matrix}}{O : \wedge\Gamma_2^* \supset A^*} \text{Rdw}}{M : \wedge\Gamma_2^* \supset A^*} (\text{O}\subseteq\text{M})_r \qquad M : \Gamma_2^*}{\dfrac{}{M : A^*}}$$

Notice that OM pairs give us the possibility to implement weaker forms of amalgamation. The "degree" of amalgamation can be tuned by varying the amalgamation of the object and the meta languages, and by changing the applicability conditions of $(\text{O}\subseteq\text{M})_r$ and $(\text{M}\subseteq\text{O})_r$.

6.4 First order OM pairs

An interesting case, which is widely studied in the literature, is that of a first order metatheory for a first order object theory. Let us see how OM pairs and the results of this paper can be applied to this case.

Definition 29 (FOL OM pair) A *First Order Object-Meta Pair* (FOL OM pair) OM is an OM pair where O and M are first order theories and L_M contains the propositional meta language of the sentences (closed formulas) of L_O.

Analogously to the propositional case, we suppose that the inference rules Δ_O and Δ_M are the set of inference rules for First Order Natural Deduction (the system C defined in [30]). Most of the results of this paper are independent on the specific inference machineries Δ_O and Δ_M. They are, therefore, applicable to FOL OM pairs. The theorems that hold also for FOL OM pairs are the following:

Theorem 5 The partial order between sets of reflection rules depicted in Figure 2

Theorem 12 The axiomatic characterization of $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ in the case of single direction reflection rules

Theorem 13 The axiomatic characterization of $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ for the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw}$, $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rup_r^n}$, $\mathrm{Rup} + \mathrm{Rdw}$, and $\mathrm{Rup} + \mathrm{Rdw} + \mathrm{Rup_r^n}$, in terms of the axioms schemata (K), (Comp), and (nTbot)

Theorem 14 The axiomatic characterization of $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ for the reflection rules $\mathrm{Rup_r^n} + \mathrm{Rdw}$ and $\mathrm{Rup^n} + \mathrm{Rdw}$

Theorem 17 The fixpoint characterization of $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ for any combination of reflection rules

Theorem 20 Reflection rules $\mathrm{Rup_r} + \mathrm{Rdw_r}$, and $\mathrm{Rup_r} + \mathrm{Rdw} + \mathrm{Rdw_r^n}$ do not have a schematic characterization

Theorem 25 The axiomatic and the fixpoint characterizations of $\mathrm{TH}_{\mathrm{OM}}(O)$ and $\mathrm{TH}_{\mathrm{OM}}(M)$ for dual OM pairs. In this case the function *dual*(.) is extended as follows: $dual(Qx.A) = Qx.dual(A)$, with $Q = \forall$ or $Q = \exists$

Theorem 26 The axiomatic characterization of dual FOL OM pairs

Theorem 27 The characterization of FOL OM pairs with empty meta theory

Notice that, all the main results proved for the propositional case are extendable to the first order case.

7 Related Work

7.1 Reflection Principles in Formal Logics

The ideas described in this paper are complementary to the large amount of work on reflection principles started by Feferman with his seminal work [12] and still under development in the field of formal logic [6]. Most of the reflection principles studied in logics are stated inside a single theory; they are variations of the following formula schema:[4]

$$\forall x \ \mathrm{P}rov(\ulcorner \phi(\dot{x}) \urcorner) \supset \phi(x) \tag{45}$$

$$\forall x \ \phi(x) \supset \mathrm{P}rov(\ulcorner \phi(\dot{x}) \urcorner) \tag{46}$$

It is well known that extending an enough expressive theory (e.g., Peano Arithmetic) with (45) and (46) leads to inconsistency. This means that any consistent theory cannot prove a full reflection principle for itself. Most of the research in the area of formal logic principles has focused on studying how to define consistent theories that satisfy some restricted versions of (45) and (46). In many cases the weakening is based on syntactic restrictions on $\phi(x)$. Examples of weakening are, *local reflection principle*, where ϕ is a closed formula, $\Sigma_n[\Pi_n]$-*reflection principle*, where ϕ is a $\Sigma_n[\Pi_n]$-formula of the arithmetical hierarchy. OM pairs constitute a new tool, which opens a

[4] Dotted variables stands for unquoted variables. More precisely, $\ulcorner \phi(\dot{x}) \urcorner$ is a term with the free variable x, such that, for any term t, the formula $x = \ulcorner t \urcorner \supset \ulcorner \phi(\dot{x}) \urcorner = \ulcorner \phi(t) \urcorner$, is valid.

new direction of investigation in reflection principles. In this line of work, reflection principles are not axiom schemata in an amalgamated language, but bridge rules between theories. Different forms of reflection principles can be obtained by selecting the different combinations of reflection rules studied in this paper. The bridge rules considered in this paper, however, corresponds to Local Reflection, as they are defined only on closed formulas. Generalized reflection principles should be formalized by reflection rules such as:

$$\frac{A(x)}{\bullet(\text{``}A(\dot{x})\text{''})}\text{ Rup} \qquad \frac{\bullet(\text{``}A(\dot{x})\text{''})}{A(x)}\text{ Rdw} \tag{47}$$

Many results proved in this paper, as for instance the partial order described in Section 3, can be easily generalized, Other results can be obtained by instantiating the general formalism for bridge rules with free variables called DFOL, developed in [13].

7.2 Metalogic Programming

Most of the approaches to Metalogic Programming (MLP) are based on the following common idea: "Given a logic program O, called object program, and a property P of this program (e.g., provability), build another program M, called the metaprogram, so that M *correctly represents* P." Examples of these approaches are described in [8, 22, 3, 25]. The approach described in [10] can be seen as a generalization of this class, as Brogi and Turini consider a Metaprogram that reasons about a set of objects programs, and about all the object programs that can be obtained by composing them. To reach this goal one has to perform three main steps: first, define the metalanguage in which it is possible to speak about the program O and the property P; second, generate the metaprogram M in the metalanguage that describes O; third, prove the fact that M correctly represents the property P of O. The statement of correctness is often referred as the *reflection principle*, and it is a sentence of the form:

The property P holds for an expression e of L_O, iff $M \vdash \phi_P(\lceil e \rceil)$.

where ϕ_P and $\lceil e \rceil$ are a formula and a term in the metalanguage that represent the property P and the expression e, respectively. In most cases P stands for provability (or derivability) in O, $\phi_P(x)$ is the unary predicate $demo(x)$, and $\lceil e \rceil$ is a closed or an open term[5] denoting the expression e. In these approaches reflection principles are statements (usually proved in the corresponding formalism) of the following form:

$$O \vdash e, \text{ if and only if } M \vdash demo(\lceil e \rceil) \tag{48}$$

$$O \vdash \neg e, \text{ if and only if } M \vdash \neg demo(\lceil e \rceil) \tag{49}$$

The analogy between OM pairs and MLP is strong. Indeed the object program and metaprogram correspond to object theory and metatheory, the encoding function $\lceil . \rceil$ corresponds to the naming function "$.$", and the reflection principles (either in the form (48) and (49), or in the form (51) and (50)) correspond to reflection rules.

[5] There are two alternative ways to represent expressions in the metalanguage: the ground representation and the non-ground representation [22]. In the first $\lceil e \rceil$ does not contains variables, in the second $\lceil e \rceil$ contains the same variables occurring in e.

However, the methodology and the final goal of MLP and of OM pairs are quite different. Indeed in MLP the starting point is an *object program* and the metaprogram *describes* some of its properties without modifying it[6] (see for instance [22] Section 4.2). In OM pairs, instead, the starting point is constituted by a partial specification of the object and the metatheory and a reflection principle formalized by a set of reflection rules; the result is the *minimal* object theory and metatheory satisfying the initial partial specifications and the reflection principle. In OM pairs, as shown by Theorem 27, when the starting metatheory is empty, the resulting metatheory describes the starting object theory without modifying it, as it happens in MLP. Going more in detail, let us compare MLP and OM pairs on the basis of their components:

Metalanguage In MLP, the metalanguage is usually composed of quantifier free first order clauses, while the OM pairs studied in this paper adopt a propositional metalanguage. This is, of course, a limitation, however, in Section 6.4 we have shown how most of the results proved for propositional OM pairs can be extended to FOL OM pairs.

Object and metaprograms (theories) In MLP, object and meta programs can be separated or amalgamated (in this latter case object and meta programs are embedded in a single logic program). In the first case there is no formal connection between the two programs. The relation between them, i.e., the reflection principles (48) and (49), is stated in the informal external theory that describes the system (see for instance Theorem 2.2.1 and Theorem 2.2.2 in [22]). In some cases reflection principles are implemented by a set of "improper" inference rules, which do not belong to a formal system, but which are shortcuts, useful for simplifying reasoning. Such rules do not add any new theorem to the system. The paradigmatic examples of such rules are the following:

$$\frac{\vdash_M \mathit{demo}(\lceil p \rceil)}{\vdash_O p} \qquad \frac{\vdash_O p}{\vdash_M \mathit{demo}(\lceil p \rceil)} \tag{52}$$

In the amalgamated approach the object program is embedded in the metaprogram. In this case it is possible to formalize the reflection principles in the amalgamated program. This is done by a set of "hybrid" clauses, which contain formulas of the object language and the metalanguage, as those shown in (50) and (51).

[6] A notably different approach, which is more in the generative spirit of OM pairs, is that by Subrahmanian, described in [33], and summarized in [29]. In this approach the metaprogram does not speak about any specific object program, so there is no necessity to prove its correctness. Metaprograms allow for the definition of metapredicates via a set of clauses. An example of such clauses are the following [33, 29]:

$$\mathit{innocent}(a) \leftarrow \neg \mathit{demo}(\lceil \mathit{guilty}(b) \rceil) \tag{50}$$

$$\mathit{demo}(V) \leftarrow V \tag{51}$$

were V is a variable that stands for any set of literals. Clauses such as (50) define an object predicate in terms of some metaproperty (provability in this case), while clauses such as (51) defines the metapredicate in terms of object properties. The goal of this approach is not to prove that the metatheory satisfies a set of reflection principles, as they are explicitly stated in the theory. The goal is rather to adapt the semantics of logic programming to the new statements like (50) and (51).

In OM pairs, object and metatheory are kept separated as in the non-amalgamated approach. Nevertheless the amalgamated approach is also representable in OM pair by the rules $(\mathrm{O}\subseteq\mathrm{M})_r$ and $(\mathrm{M}\subseteq\mathrm{O})_r$, as described in Section 6. OM pairs, however, allow the "improper" inference rules (52) to be part of the formalism. Indeed OM pair are based on Multi Context Systems [14, 31], a formal framework which allows us to cope with rules across different theories. From Multi Context System we can borrow the semantics for such rules [16], and we can compare different reflection principles in a well founded formalism, as we have done in this paper.

Metaprogram (theory) generation In the non-amalgamated approach the metatheory is defined only informally for any object program. The metaprogram generation is not part of the logic. Examples of metaprograms generated in such a way are the *Vanilla Meta Interpreter*, the *Proof Tree Meta Interpreter*, the *Instance-Demo Program* and the *SLD-Demo Program* (see [22] for a survey of these programs). For each of these metaprograms it is necessary to prove the reflection principle. This proof, however, often does not guarantee minimality, that is, the metaprogram generated is in general not the minimal metaprogram that satisfies the reflection principle. This is the case, for instance, of the *Vanilla Meta Interpreter* which is not the minimal program that satisfies the reflection principle 48. As observed in [25], in certain situations non-minimality can constitute a problem.

In OM pairs, instead, we proceed in the opposite direction: the desired reflection principles are imposed by a set of logical rules which generate the *minimal* metatheory satisfying them. OM pairs constitute a simple solution to the problem raised in [25], as the minimal metatheory satisfying (48) and (49) can be obtained simply by the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw_r}$. As shown in Section 3, this metatheory is weaker than that generated by $\mathrm{Rup_r} + \mathrm{Rdw}$ equivalent to the *Vanilla Meta Interpreter.*

In MLP there are some approaches that, due to their great similarity with OM pairs, demand a specific discussion. This is the case of the approaches which use reflection principles in a more generative way. A paradigmatic example of this class is the work of *Brogi et. al.* described in [10], and that of *Barklund et. al.* described in [4].

In [10] the metatheory formalizes provability of a set of primitive object programs in the usual way (i.e., via the Vanilla Meta Interpreter). The metatheory formalizes how composite programs can be built from primitive programs with a set of operators (i.e., $\cup$, $\cap$, etc.) An example of the axioms contained in the metatheory are:

$$demo(x \cup y, z \leftarrow w) \quad \leftarrow \quad demo(x, z \leftarrow w) \tag{53}$$

$$demo(x \cup y, z \leftarrow w) \quad \leftarrow \quad demo(y, z \leftarrow w) \tag{54}$$

In this case the reflection principle is exploited in a generative way to prove theorems in composite object programs. In terms of reflection rules, this approach can be represented with an extension of OM pairs composed of a set of "primitive" object theories, which are connected with a $\mathrm{Rup_r}$ to the metatheory, and a set of "composite" object theories which are connected to the metatheory with a Rdw. The results about OM pairs (and the other set of bridge rules) developed in this paper can be applied to this approach by generalizing composition of programs not only on the basis of provability but also on the basis of other properties such as consistency and truth.

For instance we might be interested in defining a composite program P which proves the theorems of another program P_1 and the formulas which are consistent with the program P_2.

In [4] inference rules similar to reflection rules are fully exploited for the generation of the object theory and the metatheory in the same spirit of OM pairs. Barklund introduces the concept of *theory system* as a collection of interdependent theories, some of which stand in a object/meta relation, forming an arbitrary number of metalevels. The language of theory systems contains formulas of the form:

$$\lceil t \vdash \phi \rceil$$

which stands for ϕ is provable in the theory t. In the formalism of OM pairs $\lceil t \vdash \phi \rceil$ corresponds to $\bullet(``\phi")$ in the metatheory, where the object theory term t is left implicit.

The inference machinery of theory systems is a labelled refutation system in which resolution and reduction ad absurdum can be applied inside each theory. For every theory, with names t_1 and t_2, there is a theory $t_1 \circ t_2$ that represents how the theory t_2 is seen by the theory t_1. In other words the theory t_1 acts as the metatheory for the theory $t_1 \circ t_2$. The rules which connect theories with metatheories are syntactic variations of $\mathrm{Rup_r}$ and $\mathrm{Rdw_r}$ and look as follows:

$$\mathrm{TD}\frac{t_1 \vdash \lceil t_2 \vdash \phi \rceil}{t_1 \circ t_2 \vdash \phi} \qquad \mathrm{TU}\frac{t_1 \circ t_2 \vdash \phi}{t_1 \vdash \lceil t_2 \vdash \phi \rceil}$$

In the theory t_1 it is possible to represents provability in $t_1 \circ t_2$ by the metapredicate $\lceil t_2 \vdash \dot{x} \rceil$ (where $\dot{x}$ stands for the unquoted variable x). [4] provides a semantics but does not provide a soundness and completeness result.

On the one hand it is clear that the formalism of OM pairs is representationally weaker than theory systems, since OM pairs consider only two theories. The advantages of this simple structure however is that we concentrate only on the aspects about the possible object/meta relations. On the other hand the results about OM pairs seem general enough to be extended to any pair of theories embedded in a more complex structure such as theory systems.

7.3 Weyhrauch's metatheory in FOL

In OM pairs, it is possible to formalize the reflective reasoning described by Weyhrauch in [36] and implemented in the proof checking system FOL [35, 34] and in its evolution GETFOL [15, 7]. FOL provides a way of reasoning via reflection principles between an object and a metatheory by the command REFLECT. Consider for instance the FOL command executed in the object theory:

$$\texttt{****> REFLECT MF OF}_1\texttt{ OF}_2 \tag{55}$$

where MF is the fact of the metatheory containing the theorem

$$\forall\, thm_1\ thm_2.\ \texttt{THEOREM}(mkand(thm_1, thm_2))$$

and $\texttt{OF}_1, \texttt{OF}_2$ are two facts of the object theory asserting two theorems, say ϕ_1 and ϕ_2. As described in [36, 18], the execution of command (55) generates a new fact in the object theory that is the evaluation of the function *mkand* to the arguments $\mathtt{OF}_1$ and $\mathtt{OF}_2$, i.e, a new fact $\mathtt{OF}_3$ which asserts that the conjunction $\phi_1 \wedge \phi_2$ is a theorem of the object theory. FOL reflection principles can be formalized with an OM pair containing $\mathrm{Rup_r}$ and Rdw. For instance (55) corresponds to the following deduction: $\mathrm{Rup_r} + \mathrm{Rdw_r}$.

$$\dfrac{\dfrac{O:\phi_1}{M:\bullet(``\phi_1")}\,\mathrm{Rup_r} \quad \dfrac{O:\phi_2}{M:\bullet(``\phi_2")}\,\mathrm{Rup_r} \quad M:\bullet(``\phi_1")\wedge\bullet(``\phi_2")\supset\bullet(``\phi_1\wedge\phi_2")}{\dfrac{M:\bullet(``\phi_1\wedge\phi_2")}{O:\phi_1\wedge\phi_2}\,\mathrm{Rdw_r}} \tag{56}$$

Let's consider the analogies between deduction (56) and the execution of (55) in FOL. The axioms $O:\phi_1$ and $O:\phi_2$ correspond to the facts $\mathtt{OF}_1$ and $\mathtt{OF}_2$ respectively. The axiom $M:\bullet(``\phi_1")\wedge\bullet(``\phi_2")\supset\bullet(``\phi_1\wedge\phi_2")$ corresponds to an instance of the metafact MF. Notice that, the variables $thm_1$ and $thm_2$ of sort theorem in FOL, have been substituted with the preconditions $\bullet(``\phi_1")$ and $\bullet(``\phi_2")$. The application of $\mathrm{Rup_r}$ corresponds to the bindings of the variables thm_1 and thm_2 to the formulas contained in the $\mathtt{OF}_1$ and $\mathtt{OF}_2$. The application of $\mathrm{Rdw_r}$ corresponds to the evaluation of the term $mkand(thm_1, thm_2)$, and the assertion of the results as a theorem in the object theory.

7.4 Modal logics

Amalgamated OM pairs have a tight connection with normal modal logics, under the translation of $\bullet(``A")$ into $\Box A$. Notice indeed that the proof of the modal axiom K can be obtained by the reflection rules $\mathrm{Rdw} + \mathrm{Rup_r}$ (Theorem 13), and that the necessitation rule can be obtained by a subsequent application of $(\mathrm{M}\subseteq\mathrm{O})_r$ and $\mathrm{Rup_r}$. This intuition can be easily proved.

Theorem 30 Let OMK be the amalgamated OM pair composed of the empty object theory, and the empty metatheory connected by Rdw, $\mathrm{Rup_r}$ and $M\subseteq O$. Let $(.)^*$ be the transformation on the language of the metatheory of OMK, that replaces $\bullet(``A")$ with $\Box A$, preserves the other atomic formulas, and distributes over the connectives. Then:

$$\vdash_{OMK} M:A \quad\Longleftrightarrow\quad \vdash_{\mathrm{K}} A^*$$

Proof ($\Longrightarrow$) By induction on the number of applications of $\mathrm{Rup_r}$ contained in the deductions in OMK. Any proof of $M:A$ in OMK that contains a $\mathrm{Rup_r}$ has the following form:

$$\dfrac{\dfrac{\overset{M:\Gamma}{\overset{\Pi_1}{M:\bullet(``A_1")}}}{O:A_1}\,\mathrm{Rdw} \quad \dots \quad \dfrac{\overset{M:\Gamma}{\overset{\Pi_k}{M:\bullet(``A_k")}}}{O:A_k}\,\mathrm{Rdw} \quad \dfrac{\overset{\Pi_{k+1}}{M:A_{k+1}}}{O:A_{k+1}}\,M\subseteq O \quad \dots \quad \dfrac{\overset{\Pi_n}{M:A_n}}{O:A_n}\,M\subseteq O}{\dfrac{\overset{\Pi_O}{O:A}}{\underset{\Pi}{M:\bullet(``A")}}\,\mathrm{Rup_r}}$$

where Π_O is a deduction with propositional rules of A from $A_1,\dots,A_n$, Π_i $(1\le i\le k)$ are deductions of $M:\bullet(``A_i")$ from $M:\Gamma$, and Π_j $(k+1\le j\le n)$ are proofs of $M:A_j$. By induction we have that $\vdash_K \wedge\Gamma^*\supset\Box A_i^*$ and that $\vdash_K A_j^*$. By necessitation we have that $\vdash_K \Box A_j^*$. From the fact that $A_1, \ldots A_n$ propositionally entails A, we have that $\vdash_K \Box(A_1^* \wedge \ldots \wedge A_n^* \supset A^*)$. By axiom K we can therefore conclude that $\vdash_K \wedge\Gamma \supset \Box A^*$. This concludes the only if direction.

($\Longleftarrow$) The if direction is trivial as the translation of necessitation is an admissible rule in OMK, and the translation of K is provable in OMK.

A version of Theorem 30 has been proved in [17]. In this paper, however, modal K is proved equivalent to a multi context system called MBK composed of an infinite set of theories, rather than to the amalgamated OM pair OMK. MBK is a chain-wise hierarchy of theories with a top theory. Each pair of adjacent theories constitute an OM pair where the theory above is the metatheory, the theory below is the object theory, and they are connected by the reflection rules $\mathrm{Rup_r} + \mathrm{Rdw}$.

8 Conclusion

The starting point of this paper is the notion of OM pair. OM pairs contain an object theory, a metatheory distinct from the object theory, and a set of reflection rules. Reflection rules allow for the deductive generation of the object and meta theory. This is achieved by exporting consequences from one theory to the other.

The main body of the paper consists of a study of the meta and object theories generated by the various combinations of reflection rules. We have studied the relative strength of the object and metatheories generated by different combinations of bridge rules. We have shown how our approach allows us to define metatheories already studied in the literature, but also others which have not been defined with other approaches. We have defined a duality property. Finally we have studied three important case studies: empty metatheory, Horn metatheory, amalgamated object and metatheory.

All the analysis in this paper is proof theoretical. In a following paper we will perform a semantical analysis. This will give us a new perspective on OM pairs and the formalization of metareasoning, and it will also allow us to strengthen and generalize some of the results presented here.

References

[1] G. Attardi and M. Simi. Reflections about reflection. In *Principles of Knowledge Representation and Reasoning, Proceedings of the Second International Conference*, pages 22–31. Morgan Kaufmann, 1991.

[2] G. Attardi and M. Simi. A formalisation of viewpoints. *Fundamenta Informaticae*, 23(2–4):149–174, 1995.

[3] J. Barklund. Metaprogramming in Logic. Technical Report UPMAIL Technical Report No. 80, Computing Science Department, Uppsala University, July 1994. `http://www.csd.uu.se/papers/reports.html`.

[4] J. Barklund, K. Boberg, P. Dell'Acqua, and M. Veanes. Meta-programming with Theory Systems. In *Meta-logics and Logic Programming*, chapter 8, pages 195–226. MIT Press, 1995.

[5] David Basin and Robert Constable. Metalogical frameworks. In G. Huet and G. Plotkin, editors, *Logical Environments*, pages 1–29. Cambridge University Press, Cambridge, MA, 1993. Also available as Technical Report MPI-I-92-205.

[6] Lev Beklemishev. Provability and Reflection. Lecture Notes at ESSLLI 97. Electronic copy available at `http://www.folli.uva.nl/Esslli/1997/ESSLLI97/L05/B1.PS.Z`.

[7] M. Benerecetti and L. Spalazzi. Metafol: Program tactics and logic tactics plus reflection. *Future Generation Computer Systems*, 12:139–156, 1996.

[8] K.A. Bowen and R.A. Kowalski. Amalgamating language and meta-language in logic programming. In S. Tärnlund, editor, *Logic Programming*, pages 153–173, New York, 1982. Academic Press.

[9] K.A. Bowen and T. Weiberhg. A Meta-level Extension of Prolog. In *IEEE Symposium on Logic Programming*, pages 669–675, Boston, 1985.

[10] A. Brogi and F. Turini. Meta-logic for Program Composition: Semantics Issues. In *Meta-logics and Logic Programming*, chapter 4, pages 83–110. MIT Press, 1995.

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

[12] S. Feferman. Transfinite Recursive Progressions of Axiomatic Theories. *Journal of Symbolic Logic*, 27:259–316, 1962.

[13] C. Ghidini and L. Serafini. Distributed First Order Logics. In D. Gabbay and M. de Rijke, editors, *Frontiers Of Combining Systems 2*, Studies in Logic and Computation. Research Studies Press, 1998.

[14] 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, 1993, pp. 39–49. Also IRST-Technical Report 9211-20, IRST, Trento, Italy.

[15] F. Giunchiglia. GETFOL: Interactive Multicontext Theorem Proving (abstract). In *Proceedings of IJCAI-93 Workshop on Automated Theorem Proving*, page 43, Chambery, France, 1993.

[16] F. Giunchiglia and C. Ghidini. Local Models Semantics, or Contextual Reasoning = Locality + Compatibility. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 282–289. Morgan Kaufmann, 1998. Also IRST-Technical Report 9701-07, IRST, Trento, Italy.

[17] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: how we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994. Also IRST-Technical Report 9110-07, IRST, Trento, Italy.

[18] F. Giunchiglia and A. Smaill. Reflection in constructive and non-constructive automated reasoning. In H. Abramson and M. H. Rogers, editors, *Proc. of META-88, Workshop on Metaprogramming in Logic*, pages 123–145. MIT Press, 1988. Also IRST-Technical Report 8902-04 and DAI Research Paper 375, University of Edinburgh.

[19] F. Giunchiglia and P. Traverso. A Metatheory of a Mechanized Object Theory. *Artificial Intelligence*, 80(2):197–241, 1996. Also IRST-Technical Report 9211-24, IRST, Trento, Italy, 1992.

[20] John Harrison. Metatheory and Reflection in Theorem Proving: A Survey and Critique. Technical Report CRC-053, SRI Cambridge, Millers Yard, Cambridge, UK, 1995. Available on the Web as `http://www.cl.cam.ac.uk/users/jrh/papers/reflect.dvi.gz`.

[21] P. M. Hill and J.W. Lloyd. The Gödel Programming Language. Technical Report CSTR 92-27, University of Bristol, Dept. Computer Science, 1992.

[22] P.M. Hill and J.G. Gallagher. Meta-Programming in Logic Programming. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *The Handbook of Logic in AI and Logic Programming*, volume 5, pages 421–498. Oxford University Press, 1998.

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

[24] R. Kowalski and J.S. Kim. A Metalogic programming approach to multi-agent knowledge and belief. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation - Papers in honor of John McCarthy*, pages 231–246. Academic Press, 1991.

[25] B. Martens and D. De Schreye. Two Semantics for Definite Meta-programs, using the Non-ground Representations. In *Meta-logics and Logic Programming*, chapter 3, pages 57–82. MIT Press, 1995.

[26] J. McCarthy and S. Buvač. Formalizing Context (Expanded Notes). In A. Aliseda, R.J. van Glabbeek, and D. Westerståhl, editors, *Computing Natural Language*, volume 81 of *CSLI Lecture Notes*, pages 13–50. Center for the Study of Language and Information, Stanford University, 1998.

[27] D. Perlis. On the consistency of commonsense reasoning. *Computational Intelligence*, 2:180–190, 1986.

[28] D. Perlis. Languages with Self-Reference II: Knowledge, Belief, and Modality. *Artificial Intelligence*, 34:179–212, 1988.

[29] D. Perlis and V.S Subrahmanian. Meta-language, reflection principles and self-reference. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *The Handbook of Logic in AI and Logic Programming*, volume 2, pages 339–358. Oxford University Press, 1994.

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

[31] L. Serafini and F. Giunchiglia. ML Systems: A Proof Theory for Contexts. Technical Report 0006-01, ITC-IRST, Trento, Italy, 2000. Submitted to the Journal of Logic Language and Information.

[32] C. Smorynski. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.

[33] V.S. Subrahmanian. A simple formulation of the theory of metalogic programming. In H. Abramson and M. Rogers, editors, *Meta Programming in Logic Programming*, chapter 4, pages 65–101. The MIT Press, Cambridge, MA, 1989.

[34] C. Talcott and R.W. Weyhrauch. Towards a theory of mechanizable theories: 1. fol contexts - the extensional view. In L. Carlucci Aiello, editor, *Proc. 8th European Conference on Artificial Intelligence*, pages 634–639, 1990.

[35] R.W. Weyhrauch. A Users Manual for `FOL`. Technical Report STAN-CS-77-432, Computer Science Department, Stanford University, 1977.

[36] R.W. Weyhrauch. Prolegomena to a Theory of Mechanized Formal Reasoning. *Artificial Intelligence*, 13(1):133–176, 1980.