Indistinguishability by Default

ARIEL COHEN, MICHAEL KAMINSKI, JOHANN A. MAKOWSKY

For Dov Gabbay on the occasion of his 60th birthday

ABSTRACT. We motivate and formalize the notion of *indistinguishability by default* : two objects are considered indistinguishable if they cannot be proved to be different. This idea turns out to be useful for a number of widely different applications, including natural language processing, reasoning with incomplete information, and even philosophical paradoxes.

We formalize this notion as a simple, two-element set of defaults in Reiter's default logic. We prove that the resulting default logic has a *unique* extension, in which every two objects are indistinguishable by default if and only if their non-equality cannot be proved from the known facts. We show that this relation has some desirable properties: it is reflexive, symmetric, and, although being non-transitive, it has a "transitive flavor".

All animals are equal,
But some animals are more equal than others.
George Orwell, Animal Farm ([Orwell, 1945])

1 Introduction

As far back as 1973, Dov Gabbay and Julius Moravcsik, [Gabbay and Moravcsik, 1973], raised the following question: "Is 'same' always used in the same sense? Do the truth conditions of statements of the form 'x is the same as y' remain constant throughout the variety of uses in which they may be embedded?" They go on to consider several relations, each of which, in some sense, can be interpreted as saying that x and y are the same. In this paper we concentrate on yet another "sameness" relation: *indistinguishability by default.*

The idea underlying this notion has been proposed, though not formalized, in [Charniak, 1988]:[1] two objects are considered indistinguishable unless *proved* different, in which case they are, of course, distinguishable. By

[1] Charniak calls the relation "nonmonotonic equality".

the axioms of ordinary equality, proving that t_1 and t_2 are different typically involves proving that for some formula ϕ, $\phi(t_1)$ yet $\neg\phi(t_2)$. Charniak's approach was further explored by A. Cohen and J.A. Makowsky in [Cohen and Makowsky, 1993], but the paper remained incomplete.

A considerable body of work in logic programming is based on the conceptually opposite principle, the *unique name assumption* ([Reiter, 1980a]), stating that two objects are different unless they can be proved to be equal. This fact, and the lack of formalization, may help explain why [Charniak, 1988] has not been followed up by the logic programming community. It therefore behooves us to provide some motivation for the notion of indistinguishability be default.

Such a relation may be useful for several applications, of which [Charniak, 1988] gives two examples. One is the task of resolving anaphora in natural language texts. A plausible way to do this is to reason by default — unless there is some incongruity between the pronoun and a candidate noun phrase (e.g., the pronoun is singular but the noun phrase is plural), assume that the pronoun refers to the noun phrase. Another potential use occurs in systems which attempt to understand natural language stories — in particular, when they face the task of identifying a motivation for some action described in the story. This is done by matching specific actions to stored plans. For example, suppose a story mentions a man lighting a specific match, and the program knows about a plan for lighting cigarettes, which contains the step of lighting some flammable object. The match will then be identified as the flammable object used, unless there are some facts to the contrary.

A different application where indistinguishability by default may be useful is McCarthy's Domain Circumscription ([McCarthy, 1977]): the idea that the domain should be kept as small as possible, i.e., as small as is consistent with what is provable. McCarthy argues that "humans often use circumscription, and robots must too" (see [McCarthy, 1977]). A deeper justification is given in [Davis, 1980], where the author sees circumscription as "an idea inherent in Occam's razor: only those objects should be assumed to exist which are minimally required by the context."

Yet another motivation for indistinguishability by default is more theory specific. It involves *Rough Set Theory* ([Pawlak, 1982]). The basic idea underlying Rough Set Theory is not to treat all properties as equally important: some properties are defined to be "distinguished." Two elements are considered *indiscernible* if they agree on all their distinguished properties. Relative to a given set of distinguished properties, the indiscernibility relation is clearly an equivalence relation, so it induces equivalence classes. For example, if the set of distinguished properties is $\{P, Q\}$, there are four equivalence classes: elements which satisfy $P \wedge Q$, elements which satisfy

$R \wedge \neg Q$, elements which satisfy $\neg P \wedge Q$, and elements which satisfy $\neg P \wedge \neg Q$. Rough Set Theory, which deals with equivalence classes rather than directly with the elements of the domain, turns out to be useful for a number of applications, including data mining and uncertain reasoning.

Suppose, now, that we do not have perfect information. For instance, we know that $P(t_1)$ and $Q(t_1)$, but we only know that $P(t_2)$; we do not know whether $Q(t_2)$ or $\neg Q(t_2)$. Intuitively, as far as we know, t_1 and t_2 are indiscernible. However, the resulting relation is no longer an equivalence relation, because it is not, in general, transitive ([Kryszkiewicz, 1999]).

This indiscernibility relation is, intuitively, indistinguishability by default: we decide that t_1 and t_2 are the same if this is consistent with all their known (distinguished) properties. As we will show below, indistinguishability by default has formal properties similar to uncertain indiscernibility: it is reflexive, symmetric, but not transitive. Hence, indistinguishability by default can be seen as a way to formalize the relation of rough indiscernibility in the context of imperfect information.

A very different motivation for indistinguishability by default is a philosophical one: the ancient Greek *paradox of the Sorites*. Consider a sequence of objects $t_1, t_2, \ldots, t_n$ such that every two adjacent objects are indistinguishable from one another, yet the first and last objects are quite distinct. For example, imagine a sequence of tiles such that the leftmost tile is red, the rightmost tile is pink, whereas every two adjacent tiles are so similar in color as to be indistinguishable from one another. The paradox is that, since t_1 is red but t_n is not, there appears to be some cut-off point such that tiles to its left are red, and tiles to its right are not. However, any such arbitrarily chosen point will discriminate between two adjacent tiles that are, by hypothesis, indistinguishable. Any solution to this paradox will have to incorporate an appropriate definition of indistinguishability: one that will do justice to our intuitions, yet will not lead to paradox. It is easy to see, for example, that the relevant notion of indistinguishability cannot be transitive, or we would get the paradoxical result that red tile t_1 is indistinguishable from pink tile t_n. The notion of indistinguishability by default has turned out to be very useful in formalizing an appropriate indiscriminability relation ([Cohen, 2005]).

In this paper we provide a formalization of the intuitive notion of indistinguishability by default. Earlier work ([Cohen and Makowsky, 1993]) proposes a formalization that is appropriate for closed terms, but has undesirable consequences for terms containing variables; the formalization proposed here captures the idea of indistinguishability by default in its full generality. The paper is organized as follows. In the next section we discuss first-order default logic, which will be used in our formalization. Indistinguishability by default itself is discussed in Section 3. Finally, the last section contains some concluding remarks.

2 Background: default logic

This section consists of two parts. The first part is a non-formal description of the first-order default logic, whereas the second part contains the formal definitions.

2.1 First-order default logic

Non-monotonic logics are intended to simulate the process of human reasoning by providing a formalism for deriving consistent conclusions from an incomplete description of the world.

Reiter's *default logic* ([Reiter, 1980b]) is one of the widely used non-monotonic formalisms and maybe the only non-monotonic formalism that has a clearly useful contribution to the wider field of computer science through logic programming and database theory. This logic deals with rules of inference called *defaults* which are expressions of the form

$$(1) \quad \delta(\boldsymbol{x}) = \frac{\alpha(\boldsymbol{x}) : M\beta_1(\boldsymbol{x}), \ldots, M\beta_m(\boldsymbol{x})}{\gamma(\boldsymbol{x})},$$

where $\alpha(\boldsymbol{x})$, $\beta_1(\boldsymbol{x}), \ldots, \beta_m(\boldsymbol{x})$, $m \geq 1$, and $\gamma(\boldsymbol{x})$ are formulas of first-order logic whose free variables are among $\boldsymbol{x} = x_1, \ldots, x_n$. A default is *closed* if none of $\alpha, \beta_1, \ldots, \beta_m$, and γ contains a free variable. Otherwise it is *open*. Roughly speaking, the intuitive meaning of a default is as follows. For every n-tuple of objects $\boldsymbol{t} = t_1, \ldots, t_n$, if $\alpha(\boldsymbol{t})$ is believed, and the $\beta_i(\boldsymbol{t})$s are consistent with one's beliefs, then one is permitted to deduce $\gamma(\boldsymbol{t})$ and add it to the "belief set." Thus, an open default can be thought of as a kind of "default scheme," where free variables $\boldsymbol{x}$ can be replaced by any of the theory's objects. Various examples of deduction by defaults can be found in [Reiter, 1980b].

Whereas closed defaults have been quite thoroughly investigated, not much is known about open ones. However, interesting cases of default reasoning usually deal with open defaults, because the intended use of defaults is to determine whether an object possesses a given property, rather than accepting or rejecting a "fixed statement."

It was pointed out in [Lifschitz, 1990] that when applying open defaults one must specify *all* the objects of the underlying theory. Also, it was argued in [Kaminski, 1995] that one must distinguish between objects defined explicitly (closed terms) and objects introduced implicitly (by existential formulas, say).

In this paper we use the semantical definition of extensions for open default theories proposed in [Lifschitz, 1990] and [Kaminski, 1995], where, in contrast to the syntactical definitions in [Poole, 1988] and [Reiter, 1980b], the default free variables are treated as object variables, rather than meta-variables for the closed terms of the theory. The reason for choosing a semantical definition of extensions is that, on the one hand, it provides a complete description of the theory objects, while distinguishing between explicitly and implicitly defined ones, on the other.

Since the semantical treatment of open default theories allows one to describe all the elements of the domain under consideration, it has no syntactical counterpart within the ordinary first-order default logic, unless the domain is explicitly defined by the *domain closure assumption*, i.e., the axiom

$$(2) \quad \forall x \bigvee_{i=1}^{m} x = t_i,$$

where $t_1, \dots, t_m$ are closed. Under the *domain closure assumption*, extensions can be described syntactically by extending the underlying language of default theory with an infinite set of new constant symbols and replacing each open default with the set of all its closed instances.

It was shown in [Kaminski *et al.*, 1998] that extensions for open default theories depend on the domain cardinality (cf. [Kaminski, 1997]) and that over countable[2] or finite domains, extensions for open default theories can be described syntactically in first-order logic extended with an infinitary Carnap rule of inference

$$(3) \quad \frac{\{\varphi(t)\}_{t \in \boldsymbol{T}_{\mathcal{L}}}}{\forall x \varphi(x)}.$$

Here and hereafter $\boldsymbol{T}_{\mathcal{L}}$ denotes the set of all closed terms of a language $\mathcal{L}$.[3]

2.2 Definitions

In this section we briefly recall the definitions of default theories and the Herbrand semantics of first-order logic. *In what follows we assume that language $\mathcal{L}$ of the underlying first-order logic is countable.*

2.3 Default theories

Reiter's *default logic* ([Reiter, 1980b]) deals with rules of inference called *defaults* which are expressions of the form (1).

A *default theory* is a pair (D, A), where D is a set of defaults and A is a set of first-order sentences (axioms). A default theory is *closed*, if all its

[2] In this paper, "countable" means *infinite* countable.

[3] Obviously, (2) implies (3), and, if $\boldsymbol{T}_{\mathcal{L}}$ is finite, then the converse implication also holds.

defaults are closed. However, in general, default theories are referred to as *open*.[4]

2.4 Extensions for closed default theories

Here we present the syntactical and semantical definitions of extensions for closed default theories.

Recall that closed defaults are expressions of the form

$$\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma},$$

where α, $\beta_1, \ldots, \beta_m$, $m \geq 1$, and γ are closed formulas.

DEFINITION 1. ([Reiter, 1980b, Definition 1]) Let (D, A) be a closed default theory. For any set of sentences S let $\Gamma_{(D,A)}(S)$ be the smallest set of sentences B (beliefs) that satisfies the following three properties.

D1. $A \subseteq B$.

D2. $\boldsymbol{Th}(B) = B$, i.e., B is deductively closed.

D3. If $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in B$, and $\neg\beta_1, \ldots, \neg\beta_m \notin S$, then $\gamma \in B$.

A set of sentences E is an *extension* for (D, A) if $\Gamma_{(D,A)}(E) = E$, i.e., if E is a fixed point of the operator $\Gamma_{(D,A)}$.

Next, we present a semantical definition of extension for closed default theories. Here and hereafter, for any class of interpretations W, by $\boldsymbol{Th}_{\mathcal{L}}(W)$ we mean the set of all closed formulas over $\mathcal{L}$ satisfied by all elements of W.

DEFINITION 2. ([Guerreiro and Casanova, 1990]) Let (D, A) be a closed default theory over $\mathcal{L}$. For any class of interpretations W, let $\Sigma_{(D,A)}(W)$ be the largest class V of models of A that satisfies the following condition.

If $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in \boldsymbol{Th}_{\mathcal{L}}(V)$, and $\neg\beta_1, \ldots, \neg\beta_m \notin \boldsymbol{Th}_{\mathcal{L}}(W)$, then $\gamma \in \boldsymbol{Th}_{\mathcal{L}}(V)$.

Theorem 3 below states that the definition of extensions as the theories of the fixed points of the operator Σ is equivalent to Reiter's original definition (Definition 1).

Theorem 3 below is from [Guerreiro and Casanova, 1990], see also [Marek and Truszczyński, 1993, Theorem 3.45, p. 65].

THEOREM 3. *A set of sentences E is an extension for a closed default theory (D, A) if and only if $E = \boldsymbol{Th}_{\mathcal{L}}(W)$ for some fixed point W of $\Sigma_{(D,A)}$.*

[4] That is, a closed default theory is a particular case of an open one.

2.5 Herbrand semantics of first-order logic

In this section we define Herbrand semantics of first-order logic that is the basis of the semantical approach to open default theories.

Let b be a set that contains no symbols of the underlying language $\mathcal{L}$. We denote by $\mathcal{L}_b$ the language obtained from $\mathcal{L}$ by augmenting its set of constants with all elements of b.[5] The set of all closed terms of the language $\mathcal{L}_b$ is called the *Herbrand universe* of $\mathcal{L}_b$. A *Herbrand b-interpretation* is a set of ground (closed) atomic formulas of $\mathcal{L}_b$. Note that closed formulas over $\mathcal{L}_b$ are of the form $\varphi(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are closed terms of language $\mathcal{L}_b$ and $\varphi(x_1, \ldots, x_n)$ is a formula over $\mathcal{L}$ whose free variables are among $x_1, \ldots, x_n$. The set b is called the *base* of Herbrand b-interpretation.

Let w be a Herbrand b-interpretation and let φ be a closed formula over $\mathcal{L}_b$. We say that w *satisfies* φ, denoted $w \models \varphi$, if the following holds.

- If φ is an atomic formula, then $w \models \varphi$ if and only if $\varphi \in w$;
- $w \models \varphi \supset \psi$ if and only if $w \not\models \varphi$ or $w \models \psi$;
- $w \models \neg\varphi$ if and only if $w \not\models \varphi$; and
- $w \models \forall x \varphi(x)$ if and only if for each $t \in \boldsymbol{T}_{\mathcal{L}_b}$, $w \models \varphi(t)$.

For a Herbrand b-interpretation w we define the *$\mathcal{L}$-theory* (*$\mathcal{L}_b$-theory*) of w, denoted $\boldsymbol{Th}_{\mathcal{L}}(w)$ ($\boldsymbol{Th}_{\mathcal{L}_b}(w)$), as the set of all closed formulas of $\mathcal{L}$ ($\mathcal{L}_b$) satisfied by w. For a set of Herbrand b-interpretations W we define the *$\mathcal{L}$-theory* (*$\mathcal{L}_b$-theory*) of W, denoted $\boldsymbol{Th}_{\mathcal{L}}(W)$ ($\boldsymbol{Th}_{\mathcal{L}_b}(W)$), as the set of all closed formulas of $\mathcal{L}$ ($\mathcal{L}_b$) satisfied by all elements of W. That is, $\boldsymbol{Th}_{\mathcal{L}}(W) = \bigcap\limits_{w \in W} \boldsymbol{Th}_{\mathcal{L}}(w)$ ($\boldsymbol{Th}_{\mathcal{L}_b}(W) = \bigcap\limits_{w \in W} \boldsymbol{Th}_{\mathcal{L}_b}(w)$).

Let X be a set of closed formulas over $\mathcal{L}_b$. We say that the Herbrand b-interpretation w is a *Herbrand b-model* of X, denoted by $w \models X$, if $X \subseteq \boldsymbol{Th}_{\mathcal{L}_b}(w)$. The set of all Herbrand b-models of X will be denoted by $\boldsymbol{Mod}_b(X)$. Finally, we say that X *semantically entails* a $\mathcal{L}_b$-formula φ, denoted $X \models_b \varphi$, if each Herbrand b-model of X satisfies φ.

REMARK 4. Note that the assignment to the equality relation in interpretations is a binary relation that does not have to be identity in the domain of the interpretation, but satisfies the equality first-order axioms. (Interpretations where the assignment to the equality relation is identity in the domain of the interpretation are called *normal*, see [Mendelson, 1997, p. 78] for details.) That is, the equality relation is treated as an ordinary dyadic predicate that satisfies the equality axioms.

[5] Note that if b is uncountable, then so is $\mathcal{L}_b$.

REMARK 5. It is well-known that for an infinite set of new constant symbols b, Herbrand b-interpretations are complete and sound for first-order logic. That is, for a set of formulas X over $\mathcal{L}$ and a formula φ over $\mathcal{L}$, $X \vdash \varphi$ if and only if φ is satisfied by all Herbrand b-interpretations which satisfy X. In particular, Herbrand b-interpretations with an infinite base naturally arise in the Henkin proof of the completeness theorem ([Mendelson, 1997, Lemma 2.16, p. 89]).

2.6 Extensions for open default theories

In this section, departing from Definition 2 and following [Lifschitz, 1990] and [Kaminski, 1995] we present a definition of extensions for *open* default theories. It is known from [Kaminski, 1995] (see also Remark 7 below) that for closed default theories this definition is equivalent to the original Reiter's definition (Definition 1).

We start with the intuition underlying the definition. There are two types of objects in the domain of a default theory. One type consists of the *fixed* built-in objects which belong to $\boldsymbol{T}_{\mathcal{L}}$ and must be present in any Herbrand interpretation, and the other type consist of implicitly defined *unknown* objects which may vary from one Herbrand interpretation to other, e.g., objects introduced by existentially quantified formulas. These objects generate other unknown objects by means of the function symbols of $\mathcal{L}$. Thus, it seems natural to assume that the theory domain is a Herbrand universe of the original language augmented with a set of new (unknown) objects, cf. [Lloyd, 1993, Chapter 1, §3].

The following definition of extensions for open default theories is a relativization of Definition 2 to Herbrand b-interpretations with an infinite set of new constant symbols b. The reason for passing to a semantical definition is that, in general, it is impossible to describe a Herbrand universe by means of the standard proof theory. The only exception is the cases when the theory domain is explicitly finite (see [Kaminski, 1999]), i.e., contains axiom (2).

DEFINITION 6. (Cf. [Kaminski, 1995, Definition 27]) Let b be a set of new constant symbols and let (D, A) be a default theory. For any set of Herbrand b-interpretations W let $\Delta^b_{(D,A)}(W)$ be the largest set V of Herbrand b-models of A that satisfies the following condition.

For any default $\dfrac{\alpha(\boldsymbol{x}) : M\beta_1(\boldsymbol{x}), \ldots, M\beta_m(\boldsymbol{x})}{\gamma(\boldsymbol{x})} \in D$ and any tuple $\boldsymbol{t}$ of elements of $\boldsymbol{T}_{\mathcal{L}_b}$ if $V \models \alpha(\boldsymbol{t})$ and $W \not\models \neg\beta_i(\boldsymbol{t})$, $i = 1, 2, \ldots, m$, then $V \models \gamma(\boldsymbol{t})$.

A set of $\mathcal{L}_b$ sentences E is called a *b-extension* for (D, A) if $E = \boldsymbol{Th}_{\mathcal{L}_b}(W)$ for some fixed point W of $\Delta^b_{(D,A)}$.

We will also refer to the set b as the *base* of E.

REMARK 7. It follows from the Löwnheim-Skolem theorem, see [Chang and Keisler, 1990], that, for a closed default theory (D, A) and an infinite base b, a set of sentences is a b-extension for (D, A) if and only if it is an "ordinary" Reiter's extension for (D, A).

From now on, unless we state otherwise, we deal with infinite bases, because the cardinality of a finite base b can be extracted from the b-extension, which is undesirable in the general case.

REMARK 8. Note that for two bases b and b' of different cardinality the sets of b- and b'-extensions for an open default theory do not necessarily coincide, see [Kaminski *et al.*, 1998, Example 7.1].

3 Indistinguishability by default

Here we formalize the notion of *indistinguishability by default*. The formal definition is given in Section 3.1 and its properties are listed in Section 3.2.

Below $\sim$ is a binary predicate symbol that does not belong to the underlying language $\mathcal{L}$. It is intended to denote "indistinguishability by default."

3.1 Relation $\sim$

A two element set of defaults $\left\{ \frac{: Mx = y}{x \sim y}, \frac{x \neq y :}{x \not\sim y} \right\}$ will be denoted $D_{\sim}$.

REMARK 9. In the modal context, the intuitive meaning of $\sim$ is that $x \sim y$ if and only if "possibly $x = y$," cf. [Kaminski and Rey, 2000, Section 7].

REMARK 10. Let E be an extension for default theory $(D_{\sim}, A)$. Then E "is complete for $\sim$" in the following sense. For each two closed terms t_1 and t_2, either $t_1 \sim t_2$ or $t_1 \not\sim t_2$ belongs to E.

Let us now demonstrate that, for a given Herbrand base, the definition of'$\sim$' "works", in the sense that two terms are indistinguishable by default iff they cannot be proved to be different.

THEOREM 11. *Operator $\Delta^b_{(D_{\sim}, A)}$ has a unique fixed point*

$$(4) \quad \begin{array}{rll} V_b = \boldsymbol{Mod}_b(A & \cup & \{t_1 \sim t_2 : t_1, t_2 \in \boldsymbol{T}_{\mathcal{L}_b} \ \ \textit{and} \ \ A \not\models_b t_1 \neq t_2\} \\ & \cup & \{t_1 \not\sim t_2 : t_1, t_2 \in \boldsymbol{T}_{\mathcal{L}_b} \ \ \textit{and} \ \ A \models_b t_1 \neq t_2\}). \end{array}$$

For the proof of Theorem 11 we shall need the following notation. For a Herbrand b-interpretation w of $\mathcal{L} \cup \{\sim\}$, we denote by w^- the restriction of w to $\mathcal{L}_b$, and for a set W of Herbrand b-interpretation of $\mathcal{L} \cup \{\sim\}$, we define a set W^- of Herbrand b-interpretations of $\mathcal{L}$ by $W^- = \{w^- : w \in W\}$.

Proof of Theorem 11. Let W be a set of Herbrand b-interpretations of $\mathcal{L} \cup \{\sim\}$ satisfying A. Then $W \not\models t_1 \neq t_2$ implies $A \not\models_b t_1 \neq t_2$, which, in turn, implies $V_b \models t_1 \sim t_2$. In addition, if $V_b \models_b t_1 \neq t_2$, then $A \models_b t_1 \neq t_2$, implying $V_b \models_b t_1 \not\sim t_2$. This together with Remark 10 and (4) (in Theorem 11) implies $\Delta^b_{(D_\sim, A)}(W) = V_b$, and the theorem follows. ■

COROLLARY 12. *For each infinite set of new constant symbols b, default theory $(D_\sim, A)$ has a unique b-extension $E_b = \boldsymbol{Th}_{\mathcal{L}}(V_b)$.*

Theorem 11 and its corollary seemingly leave open the undesirable possibility that indistinguishability by default depends on the particular choice of the Herbrand base b. However, this is only an apparent possibility. Theorem 13 below shows that restrictions of b-extensions for $(D_\sim, A)$ to $\mathcal{L}$ do not depend on the cardinality of base b, cf. [Kaminski *et al.*, 1998, Example 7.1].

THEOREM 13. *Let b_1 and b_2 be infinite sets of new constant symbols and let V_{b_1} and V_{b_2} be the fixed points of $\Delta^{b_1}_{(D_\sim, A)}$ and $\Delta^{b_2}_{(D_\sim, A)}$, respectively. Then $\boldsymbol{Th}_{\mathcal{L}}(V_{b_1}) = \boldsymbol{Th}_{\mathcal{L}}(V_{b_2})$.*

The proof of Theorem 13 is based on the following observation.

LEMMA 14. *Let b_1 and b_2 be infinite sets of new constant symbols such that $b_1 \subseteq b_2$ and let φ be a closed formula over $\mathcal{L}_{b_1}$. Then $A \models_{b_1} \varphi$ if and only if $A \models_{b_2} \varphi$.*

Proof. We start with the proof of the "only if" part of the lemma. Let $A \models_{b_1} \varphi$ and assume to the contrary that $A \not\models_{b_2} \varphi$. Let w_2 be a Herbrand b_2-model of A such that $w_2 \not\models \varphi$. Let b' be the set of new constant symbols which appear in φ and let w_1 be an elementary subinterpretation of w_2 of cardinality $||b_1||$ containing b'.[6] In particular, w_1 is a Herbrand b'_1-interpretation for some subset b'_1 of b_2 of cardinality $||b_1||$. Renaming the elements of $b'_1 \setminus b'$, if necessary, we may assume that $b'_1 = b_1$. Then, being an elementary subinterpretation of w_2, $w_1 \models A \cup \{\neg\varphi\}$, which contradicts $A \models_{b_1} \varphi$.

For the proof of the "if" part of the lemma, let $A \models_{b_2} \varphi$ and assume to the contrary that $A \not\models_{b_1} \varphi$. Let w_1 be a Herbrand b_1-model of A such that $w_1 \not\models \varphi$. Let b' be the set of new constant symbols which appear in φ and let $\iota : b_2 \to b_1$ be a surjective mapping that is an identity on b_1. Consider a Herbrand b_2-interpretation

$$w_2 = \{P(c_1, c_2, \ldots, c_n) : P(\iota(c_1), \iota(c_2), \ldots, \iota(c_n)) \in w_1\}.$$

[6] Recall that $\mathcal{L}$ is countable.

A simple induction on the complexity of an $\mathcal{L}$ formula $\psi(x_1, x_2, \ldots, x_n)$ shows that for any $c_1, c_2, \ldots, c_n \in b_2$ $w_2 \models \psi(c_1, c_2, \ldots, c_n)$ if and only if $w_1 \models \psi(\iota(c_1), \iota(c_2), \ldots, \iota(c_n))$.[7] Thus, w_2 is a Herbrand b_2-model of A that does not satisfy φ, which contradicts $A \models_{b_2} \varphi$. ■

COROLLARY 15. *Let b_1 and b_2 be infinite sets of new constant symbols such that $b_1 \subseteq b_2$. Let $w_2 \in V_{b_2}$ and let w_1 be the Herbrand b_1-interpretation constructed in the proof of the "only if" part of Lemma 14. Then $w_1 \in V_{b_1}$.*[8]

Proof. Since w_1 is an elementary subinterpretation of w_2, $w_1 \models A$.

Let $t_1, t_2 \in \boldsymbol{T}_{\mathcal{L}_{b_1}}$ and assume that $A \not\models_{b_1} t_1 \neq t_2$. Then, by Lemma 14, $A \not\models_{b_2} t_1 \neq t_2$ either, which together with $w_2 \in V_{b_2}$ implies $w_2 \models t_1 \sim t_2$. Since w_1 is an elementary subinterpretation of w_2, $w_1 \models t_1 \sim t_2$ as well.

Similarly, $A \models_{b_1} t_1 \neq t_2$ implies $w_1 \models t_1 \not\sim t_2$, and $w_1 \in V_{b_1}$ follows. ■

COROLLARY 16. *Let b_1 and b_2 be infinite sets of new constant symbols such that $b_1 \subseteq b_2$. Let $w_1 \in V_{b_1}$ and let w_2 be the Herbrand b_1-interpretation constructed in the proof of the "if" part of Lemma 14. Then $w_2 \in V_{b_2}$.*

Proof. The proof is similar to that of Corollary 15.

Since w_1 is an elementary subinterpretation of w_2 and $w_1 \in w_2$, $w_2 \models A$.

Let $t_1, t_2 \in \boldsymbol{T}_{\mathcal{L}_{b_2}}$ and assume that $A \not\models_{b_2} t_1 \neq t_2$. Then $A \not\models_{b_2} \iota(t_1) \neq \iota(t_2)$.[9] Consequently, $w_1 \models \iota(t_1) \sim \iota(t_2)$, implying $w_2 \models t_1 \sim t_2$.

Similarly, $A \models_{b_2} t_1 \neq t_2$ implies $w_2 \models t_1 \not\sim t_2$, and $w_2 \in V_{b_2}$ follows. ■

Proof of Theorem 13. Of course, we may assume that $b_1 \subseteq b_2$.

Let φ be a closed formula over $\mathcal{L}$. Assume that $\varphi \notin \boldsymbol{Th}_{\mathcal{L}}(V_{b_2})$. Let w_2 be an element of V_{b_2} that does not satisfy φ and let w_1 be an elementary subinterpretation of w_2 constructed in the proof of the "only if" part of Lemma 14. Then $w_1 \not\models \varphi$ and, by Corollary 15, $w_1 \in V_{b_1}$. Thus, $\varphi \notin \boldsymbol{Th}_{\mathcal{L}}(V_{b_1})$ either.

Conversely, assume that $\varphi \notin \boldsymbol{Th}_{\mathcal{L}}(V_{b_1})$. Let w_1 be an element of V_{b_1} that does not satisfy φ and let w_2 be the Herbrand b_2-interpretation constructed in the proof of the "if" part of Lemma 14. Then $w_2 \not\models \varphi$ and, by Corollary 16, $w_2 \in V_{b_2}$. Thus, $\varphi \notin \boldsymbol{Th}_{\mathcal{L}}(V_{b_2})$, which completes the proof. ■

By Theorems 11 and 13, all restrictions of b-extensions for $(D_\sim, A)$ to $\mathcal{L}$ coincide, i.e., $\sim$ is well defined. So, regardless of the Herbrand base, $(D_\sim, A)$ has a unique extension, which we shall denote E_A.

[7] In particular, w_1 is an elementary subinterpretation of w_2.

[8] See (4) (in Theorem 11) for the definition of V_b.

[9] As usual, for a term $t \in \boldsymbol{T}_{\mathcal{L}_{b_2}}$, $\iota(t)$ is obtained from t by each $c \in b_2$ with $\iota(c)$. We leave the formal definition to the reader.

3.2 Properties of $\sim$

In this section we observe a number of properties of $\sim$ which follows easily from its definition. First we note that indistinguishability by default complies with the intuitive notion described in the introduction. That is, that two domain elements are indistinguishable if and only if there is no property that can be proved to be true of one yet false of the other. In particular, if t_1 and t_2 are closed terms, then $t_1 \sim t_2$ if and only if there exists no formula $\varphi(x)$ with only one free variable x such that $A \vdash \varphi(t_1)$, but $A \vdash \neg\varphi(t_2)$.

Below (in addition to Remark 10) we list some properties of $\sim$ which immediately follow from the definition of E_A.

- Indistinguishability by default is *non-monotonic*. For example, obviously, $\forall x \forall y x \sim y \in E_\emptyset$, However, $\forall x \forall y x \sim y \notin E_{c_1 \neq c_2}$, because of the *instance* $\frac{c_1 \neq c_2 :}{c_1 \not\sim c_2}$ of default $\frac{x \neq y :}{x \not\sim y}$, see Theorem 11.
- By definition, $\sim$ is reflexive and symmetric. However, it is not transitive. For example, both $c_1 \sim c_2$ and $c_2 \sim c_3$ belong to $E_{c_1 \neq c_3}$, whereas $c_1 \sim c_3$ does not.
- On the other hand $\sim$ possesses the following "semi-transitivity" property, which looks almost like a "duble negation" of the usual transitivity of equality.

  Let $t_1, t_2, t_3 \in \boldsymbol{T}_{\mathcal{L}}$. *Then* $A \not\vdash t_1 \neq t_2 \vee t_2 \neq t_3$ *implies* $t_1 \sim t_3 \in E_a$.[10]

  For the proof assume $A \not\vdash t_1 \neq t_2 \vee t_2 \neq t_3$. By Theorem 11, it suffices to show that $A \not\vdash t_1 \neq t_3$. Were $A \vdash t_1 \neq t_3$, by transitivity of $=$, we would have $A \vdash t_1 \neq t_2 \vee t_2 \neq t_3$, which contradicts our assumption.

Finally we show that extensions for $(D_\sim, A)$ have a "generic" element given by Proposition 17 below.

PROPOSITION 17. *Let* V_b *be the fixed point of* $\Delta^b_{(D_\sim, A)}$. *Then* $V_b \models \exists x \forall y (x \sim y)$.

Proof. Let $c \in b$ be a new constant symbol. Then for each $t \in \boldsymbol{T}_{\mathcal{L}_b}$ there is a Herbrand b-interpretation $w \in V_b$ such that $w \models c = t$. Thus, $V_b \models \forall x (c \sim x)$, and the proposition follows. ■

4 Conclusion

In this paper we have formalized the idea of considering two objects indistinguishable unless there is some property that distinguishes between them.

[10] Obviously, this property would hold, were $\sim$ transitive.

We have shown that our formalization has some desirable properties. The relation is reflexive, symmetric, and has the property of semi-transitivity. We believe, therefore, that our formalization does justice to the intuitive notion of indistinguishability by default.

BIBLIOGRAPHY

[Chang and Keisler, 1990] C.C. Chang and H.J. Keisler. *Model Theory*. North Holland Publishing Co., Amsterdam, 1990.

[Charniak, 1988] E. Charniak. Motivation analysis, abductive unification and nonmonotonic equality. *Artificial Intelligence*, 34:275–295, 1988.

[Cohen and Makowsky, 1993] A. Cohen and J. A. Makowsky. Two approaches to nonmonotonic equality. Technical Report CIS-9317, Technion—Israel Institute of Technology, 1993.

[Cohen, 2005] A. Cohen. Vagueness and indiscriminability by failure. Presented at *the 33rd Annual Meeting of the Society for Exact Philosophy*, 2005.

[Davis, 1980] M. Davis. The mathematics of non-monotonic reasoning. *Artificial Intelligence*, 13:73–80, 1980.

[Gabbay and Moravcsik, 1973] D. Gabbay and J. M. E. Moravcsik. Sameness and individuation. *Journal of Philosophy*, 70:16, 1973. Reprinted in F. J. Pelletier (ed.), *Mass Terms: Some Philosophical Problems*, Dordrecht, 1979. Reidel.

[Guerreiro and Casanova, 1990] R. Guerreiro and M. Casanova. An alternative semantics for default logic. Presented at *the 3rd International Workshop on Nonmonotonic Reasoning*, 1990.

[Kaminski and Rey, 2000] M. Kaminski and G. Rey. First–order non-monotonic modal logics. *Fundamenta Informaticae*, 42:303–333, 2000.

[Kaminski *et al.*, 1998] M. Kaminski, J.A. Makowsky, and M. Tiomkin. Extensions for open default theories via the domain closure assumption. *Journal of Logic and Computation*, 8:169–187, 1998.

[Kaminski, 1995] M. Kaminski. A comparative study of open default theories. *Artificial Intelligence*, 77:285–319, 1995.

[Kaminski, 1997] M. Kaminski. A note on the stable model semantics for logic programs. *Artificial Intelligence*, 96:467–479, 1997.

[Kaminski, 1999] M. Kaminski. Open default theories over closed domains. *Logic Journal of the IGPL*, 7:577–589, 1999.

[Kryszkiewicz, 1999] M. Kryszkiewicz. Rules in incomplete information systems. *Information Sciences*, 113:271–292, 1999.

[Lifschitz, 1990] V. Lifschitz. On open defaults. In J.W. Lloyd, editor, *Computational Logic: Symposium Proceedings*, pages 80–95, Berlin, 1990. Springer–Verlag.

[Lloyd, 1993] J.W. Lloyd. *Foundation of logic programming, second extended edition*. Springer–Verlag, Berlin, 1993.

[Marek and Truszczyński, 1993] W. Marek and M. Truszczyński. *Nonmonotonic Logic*. Springer–Verlag, Berlin, 1993.

[McCarthy, 1977] J. McCarthy. Epistemological problems of artificial intelligence. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 1038–1044, Los Angeles, CA, 1977. Kaufmann.

[Mendelson, 1997] E. Mendelson. *Introduction to mathematical logic*. Chapman and Hall, London, 1997.

[Orwell, 1945] G. Orwell. *Animal Farm*. Penguin, London, 1945.

[Pawlak, 1982] Z. Pawlak. Rough set. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.

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

[Reiter, 1980a] R. Reiter. Equality and domain closure in first order databases. *Journal of the ACM*, 27:235–249, 1980.

[Reiter, 1980b] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

INDEX