DISTANCES BETWEEN FORMAL THEORIES

MICHÈLE FRIEND, MOHAMED KHALED, KOEN LEFEVER AND GERGELY SZÉKELY

Abstract. In the literature, there have been several methods and definitions for working out if two theories are "equivalent" (essentially the same) or not. In this article, we do something subtler. We provide means to measure distances (and explore connections) between formal theories. We define two main notions for such distances. A natural definition is that of *axiomatic distance*, but we argue that it might be of limited interest. The more interesting and widely applicable definition is that of *conceptual distance* which measures the minimum number of concepts that separate two theories. For instance, we use conceptual distance to show that relativistic and classical kinematics are distinguished by one concept only. We also develop further notions of distance, and we include a number of suggestions for applying and extending our project. We end with a philosophical discussion of the significance of these approaches.

1. Introduction

It is well known that the theory of strict partial orders and the theory of partial orders are "equivalent", i.e., they have the same essential content. To capture this vague idea, defining a precise equivalence between theories in general, several formal definitions have been made, e.g., *logical equivalence*, *definitional equivalence*, *categorical equivalence*, etc. Which theory is equivalent to which other theory depends on the point of view from which one decides to explore the equivalence between the theories in question.

In the last few decades, the concept of the equivalence between theories (henceforth: "theory-equivalence") has become important for studying the connections between formal theories. Many interesting results have been derived from investigating such equivalence, cf., e.g., [Andréka et al., 2005], [Barrett and Halvorson, 2016], [Japaridze and Jongh, 1998], [Pinter, 1978] and [Visser, 2006]. We can also look at the question starting from non-equivalence. Given two non-equivalent theories (according to any chosen definition of theory-equivalence), some natural questions arise: (1) Can these theories be modified into equivalent theories (in a non-trivial way)? (2) If this can be done, can we do it in finitely many steps? In other words, what is the degree of their non-equivalence?

In the present paper, we lay down the first steps of a research programme to answer these questions. In order to investigate some ways to measure how far two theories are from each other; we introduce a framework that can give a qualitative and quantitative analysis of the connections between formal theories. We focus on first order theories, but it is worth mentioning that the same work can be extended to all formal theories of the same type. We develop several notions for distances between theories, we discuss these notions and we make comparisons between them.

2010 *Mathematics Subject Classification.* Primary 03B99, 03C07, 03A10. Secondary 03G99, 03B80.

Key words and phrases. Network of theories, degrees of non-equivalence, conceptual distance, relativistic and classical kinematics.

The idea is very simple: based on a symmetric relation capturing a notion of minimal change, we introduce a general way to define a distance on any class of objects (not just theories) equipped with an equivalence relation. The idea is a generalization of the distance between any two nodes in the same graph, in graph theory. After, we give particular examples when the given class is a class of first order theories and the equivalence relation is a fixed notion of theory-equivalence.

The first particular example, is that of logical equivalence. As a measure for the degree of logical non-equivalence, we introduce the concept of *axiomatic distance*. The idea is to count the minimum number of axioms that needed to be added or "removed" to get from one theory to the other[1]. Since any finite number of axioms can be concatenated by conjunction resulting in only one axiom, one may think that the axiomatic distance, if it is finite, between two given theories T and T' must be ≤ 2, i.e., we need at most two steps to get T from T': one step for axiom addition and another one for axiom removal. This is why we have the intuition that axiomatic distance is not very interesting, c.f., Problem 1 and Theorem 3.7 herein.

Then we turn to definitional equivalence. Two theories are definitionally equivalent if they cannot be distinguished by a concept (a formula defining some notion). As a measure for the degree of definitional non-equivalence, we define *conceptual distance*. This distance counts the minimum number of *concepts* that separate two theories. We find that this distance is of special interest in the study of logic. We give examples and we count conceptual distance between some specific theories, see, e.g., Theorem 4.10. We also explore a connection between conceptual distance and *spectrum of theories* which is a central topic in model theory, c.f., Proposition 4.13.

Furthermore, we investigate the possible application of conceptual distance in the logical foundation of physical theories in first order logic. This is indeed an interesting result in its own right, not only for logicians but also for physicists. We prove that the conceptual distance between classical and relativistic kinematics is one. In other words, only one concept distinguishes classical and relativistic kinematics: the existence of a class of observers who are at absolute rest. Such a result opens several similar questions about how many concepts (and what are they) differentiate two physical theories.

In philosophy of physics, this will be important because, on the one hand it is clear that we are not presently converging towards one unified theory of physics in the sense of converging to one set of laws from which all the phenomena of physics can be derived. On the other hand, we would like to know the exact relationship between theories. If we know this, then we can form an impression of how far we are from such a philosophical dream – the dream of the unity of physics. Or, we can adjust our hopes and expectations, and rest content with a unity of science at a more general level: as a network of logical theories with precise relations between them.

With the definitions and metrics on distance developed here, we have maps of the network of logical theories. When we draw such maps of networks, the topology may suggest very interesting and fruitful questions. For instance: if there is a distance other than zero or one, then is there already a known theory in between, or if not, we can ask what are the limitative properties of that theory and what is its philosophical significance? By engaging in such studies, we see the "edge" of the limitative results, and by examining this edge we more precisely understand the rapport between meta-logical limitative results and physical phenomena.

---

[1]By "removing an axiom" here we only mean the 'inverse' of adding an axiom in the following sense: T is a theory resulting from "removing" one axiom from T' if T' can be reached from T by adding one axiom.

2. Elements of first order logic (FOL)

In this section, we briefly recall the basic definitions and notions of first order logic. A ***language*** $\mathcal{L}$ is a set of relation symbols each of which is assigned a positive rank. Generally, a language is assumed to contain some function symbols and/or constant symbols, but every such symbol can be viewed as a relation symbol in a natural way. Throughout, unless otherwise is stated, languages, models, theories, etc, are understood to be first order languages, models, theories, etc.

2.1. **The syntax of FOL.** Given a language $\mathcal{L}$. To construct the formulas, we also need some other symbols: fixed countable infinite set of individual variables $\{v_0, v_1, \ldots\}$, equality symbol "$=$" (we deal with first order logic with identity), brackets "(" and ")", comma ",", conjunction "$\wedge$", negation "$\neg$" and the existential quantifier "$\exists$". *The set of* ***formulas*** Fm *of* $\mathcal{L}$ is the smallest set that satisfies:

(a) Fm contains each ***basic formula*** *of* $\mathcal{L}$, where the basic formulas are the following two types of formulas:
  (i) The equalities $v_i = v_j$, for any $i, j \in \mathbb{N}$.
  (ii) $R(v_{i_0}, \ldots, v_{i_m})$, for any relation symbol R of rank $m+1$.
(b) Fm contains $\varphi \wedge \psi$, $\neg\varphi$ and $\exists v_i\ \varphi$, for each $\varphi, \psi \in \mathsf{Fm}$.

Throughout, we will use metavariables, that means we use variable symbols as $x, y, z, x_0, x_1, \ldots$, etc. The convention is that a metavariable is to be uniformly substituted with the same instance from the individual variables in all its appearances in a given formula.

Definition 2.1. Let $\mathcal{L}$ be a language and suppose that Fm is the set of formulas of $\mathcal{L}$. Let $X \subseteq \mathsf{Fm}$. We define $\mathfrak{F}(X)$ to be the smallest set that satisfies the following:

(a) $X \subseteq \mathfrak{F}(X)$, and
(b) $\mathfrak{F}(X)$ contains $\varphi \wedge \psi$, $\neg\varphi$ and $\exists x\ \varphi$, for each $\varphi, \psi \in \mathfrak{F}(X)$.

In particular, one can easily see that $\mathsf{Fm} = \mathfrak{F}(\mathsf{Basics})$, where Basics is the set of all basic formulas of $\mathcal{L}$. As usual, we use the following abbreviations.

- Disjunction, implication, equivalence and universal quantifier:

$$\varphi \vee \psi \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \neg\psi) \qquad \varphi \to \psi \stackrel{\text{def}}{=} \neg(\varphi \wedge \neg\psi)$$
$$\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi \to \psi) \wedge (\psi \to \varphi) \qquad \forall x\ \varphi \stackrel{\text{def}}{=} \neg(\exists x\ \neg\varphi)$$

- We write $\bar{x}$ to mean a finite sequence of variables $\bar{x} = x_1, \ldots, x_m$. We also use the polyadic quantifiers $\exists x_1 \cdots x_m\ \varphi \stackrel{\text{def}}{=} \exists x_1\ \cdots\ \exists x_m\ \varphi$.
- We use grouped conjunction $\bigwedge$ and grouped disjunction $\bigvee$. Formally, let Σ be a finite set of formulas. If $\Sigma = \emptyset$, then $\bigvee\Sigma$ is defined to be $\neg(x = x)$ while $\bigwedge\Sigma$ is defined to be $x = x$. Otherwise, fix an enumeration of Σ, say, $\varphi_1, \ldots, \varphi_m$. Then[2]

$$\bigvee\Sigma \stackrel{\text{def}}{=} \varphi_1 \vee \cdots \vee \varphi_m \quad \text{and} \quad \bigwedge\Sigma \stackrel{\text{def}}{=} \varphi_1 \wedge \cdots \wedge \varphi_m.$$

[2] There is an ambiguity in this definition because it depends on an enumeration. Since conjunctions and disjunctions are both commutative in FOL, no real harm can be caused by this ambiguity.

2.2. **The semantics of FOL.** A ***model*** $\mathfrak{M}$ *for language* $\mathcal{L}$ is a set M enriched with operations $R^{\mathfrak{M}} \subseteq M^{m+1}$, for each relation symbol R with rank $m+1$. An ***assignment*** *in* $\mathfrak{M}$ is a function τ that assigns for each variable an element of the set M. Let $\varphi \in \mathsf{Fm}$ be any formula. The satisfiability relation $\mathfrak{M} \models \varphi[\tau]$ is defined recursively as follows:

$$\begin{aligned} \mathfrak{M}, \tau \models R(x_0, \ldots, x_m) \quad &\text{iff} \quad (\tau(x_0), \ldots, \tau(x_m)) \in R^{\mathfrak{M}} \\ \mathfrak{M}, \tau \models x = y \quad &\text{iff} \quad \tau(x) = \tau(y) \\ \mathfrak{M}, \tau \models \varphi \wedge \psi \quad &\text{iff} \quad \mathfrak{M}, \tau \models \varphi \text{ and } \mathfrak{M}, \tau \models \psi \\ \mathfrak{M}, \tau \models \neg\varphi \quad &\text{iff} \quad \mathfrak{M}, \tau \not\models \varphi \\ \mathfrak{M}, \tau \models \exists x\ \varphi \quad &\text{iff} \quad \text{there is } a \in M \text{ such that } \mathfrak{M}, \tau[x \mapsto a] \models \varphi, \end{aligned}$$

where $\tau[x \mapsto a]$ is the assignment which agrees with τ on every variable except that $\tau[x \mapsto a](x) = a$. A formula φ is said to be ***satisfiable*** *in* $\mathfrak{M}$ iff there is an assignment τ in $\mathfrak{M}$ such that $\mathfrak{M}, \tau \models \varphi$. A formula φ is said to be ***true*** *in* $\mathfrak{M}$, in symbols $\mathfrak{M} \models \varphi$, iff $\mathfrak{M}, \tau \models \varphi$, for every assignment τ in $\mathfrak{M}$. The ***theory*** *of* $\mathfrak{M}$ is defined as follows:

$$\mathsf{Th}(\mathfrak{M}) \stackrel{\text{def}}{=} \{\varphi \in \mathsf{Fm} : \mathfrak{M} \models \varphi\}.$$

2.3. **First order theories.**

Definition 2.2. Suppose that $\mathcal{L}$ is a language and let Fm be its set of formulas. A ***theory*** T *of* $\mathcal{L}$ is a set of formulas (subset of Fm).

We use the same superscripts and subscripts for theories and their corresponding languages and formulas. For example, if we write T' is a theory, then we understand that T' is a theory of language $\mathcal{L}'$ whose set of formulas is Fm'.

Let T be a theory of language $\mathcal{L}$. A ***model*** *for* T is a model for $\mathcal{L}$ in which every $\psi \in T$ is true. We say that T *is a* ***consistent*** iff there is at least one model for T. Let $\varphi \in \mathsf{Fm}$. We say that φ is ***satisfiable*** *in* T iff φ is satisfiable in some model for T. We say that φ is a ***theorem*** *of* T, in symbols $T \models \varphi$, iff φ is true in every model for T. We say that φ is a ***contradiction*** *to* T iff $T \models \neg\varphi$. The *set of* ***consequences*** *of theory* T is defined as follows:

$$\mathsf{Cn}(T) \stackrel{\text{def}}{=} \{\varphi \in \mathsf{Fm} : T \models \varphi\}.$$

Definition 2.3. Two theories T_1 and T_2 are called ***logically equivalent***, in symbols $T_1 \equiv T_2$, iff they have the same consequences, i.e., $\mathsf{Cn}(T_1) = \mathsf{Cn}(T_2)$.

2.4. **More notions for theory-equivalence.** Recall the notions of bounded and free variables in a given formula. Roughly, a variable x that occurs in a formula φ in the scope of a quantifier is said to be a ***bounded variable*** *in* φ. A variable x that occurs in a formula φ but not in the scope of any quantifier is called a ***free variable*** *in* φ. The notion of free variables specifies places in a formula where substitution may take place. We write $\varphi(x_0, \ldots, x_m)$ to mean that the free variables of φ are among $x_0, \ldots, x_m$.

Definition 2.4. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two languages. A ***translation*** *of* $\mathcal{L}_1$ *into* $\mathcal{L}_2$ is a map $\mathsf{tr} : \mathsf{Fm}_1 \to \mathsf{Fm}_2$ such that, for each formula $\varphi \in \mathsf{Fm}_1$, $\mathsf{tr}(\varphi)$ is defined by the following induction on the complexity of φ:

- $\mathsf{tr}(x = y)$ is $x = y$.

- For each relation symbol $R \in \mathcal{L}_1$ of rank $m+1$, a formula $\varphi_R(v_0, \dots, v_m)$ in Fm_2 is assigned. Then $\mathsf{tr}(R(v_{i_0}, \dots, v_{i_m}))$ is an appropriate substitution of the formula φ_R.[3]
- tr commutes with the Boolean connectives:
$$\mathsf{tr}(\neg\varphi) = \neg\mathsf{tr}(\varphi) \quad \text{and} \quad \mathsf{tr}(\varphi \wedge \psi) = \mathsf{tr}(\varphi) \wedge \mathsf{tr}(\psi).$$
- Finally, $\mathsf{tr}(\exists x\ \varphi) = \exists x\ \mathsf{tr}(\varphi)$.

Definition 2.5. Suppose that T_1 and T_2 are theories in languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, and tr is a translation of $\mathcal{L}_1$ into $\mathcal{L}_2$. The translation tr is said to be an ***interpretation*** *of* T_1 *into* T_2 iff it maps theorems of T_1 into theorems of T_2, i.e., for each formula $\varphi \in \mathsf{Fm}_1$,
$$T_1 \models \varphi \implies T_2 \models \mathsf{tr}(\varphi).$$

(a) An interpretation tr of T_1 into T_2 is called ***faithful interpretation*** iff for each formula $\varphi \in \mathsf{Fm}_1$,
$$T_1 \models \varphi \iff T_2 \models \mathsf{tr}(\varphi).$$

(b) An interpretation tr_{12} of T_1 into T_2 is called ***definitional equivalence*** iff there is an interpretation tr_{21} of T_2 into T_1 such that
- $T_1 \models \mathsf{tr}_{21}\big(\mathsf{tr}_{12}(\varphi)\big) \leftrightarrow \varphi$,
- $T_2 \models \mathsf{tr}_{12}\big(\mathsf{tr}_{21}(\psi)\big) \leftrightarrow \psi$.

for every $\varphi \in \mathsf{Fm}_1$ and $\psi \in \mathsf{Fm}_2$. In this case, tr_{21} is also a definitional equivalence.

Definition 2.6. Two theories T_1 and T_2 are said to be ***definitionally equivalent***, in symbols $T_1 \rightleftarrows T_2$, iff there is a definitional equivalence between them.

There are several ways to define definitional equivalence. We here use a variant of the definition in [Henkin et al., 1985, Definition 4.3.42]. For a discussion on the different definitions of definitional equivalence, see [Lefever and Székely, 2018b], and we refer to [Visser, 2006] for a category theory based discussion.

Proposition 2.7. Let T_1 and T_2 be two theories and suppose that tr_{12} is a definitional equivalence between T_1 and T_2, then tr_{12} is also a faithful interpretation.

Proof. Let T_1 and T_2 be two theories, and let tr_{12} be a definitional equivalence between them. Let $\varphi \in \mathsf{Fm}_1$, we should show that $T_1 \models \varphi$ iff $T_2 \models \mathsf{tr}_{12}(\varphi)$. Since tr_{12} is an interpretation, we have that $T_1 \models \varphi$ implies $T_2 \models \mathsf{tr}_{12}(\varphi)$. To show the converse, let us assume that $T_2 \models \mathsf{tr}_{12}(\varphi)$. By Definition 2.5, there is an interpretation tr_{21} of T_2 into T_1 such that $T_1 \models \mathsf{tr}_{21}\big(\mathsf{tr}_{12}(\varphi)\big) \leftrightarrow \varphi$. Since tr_{21} is an interpretation and $T_2 \models \mathsf{tr}_{12}(\varphi)$, we have $T_1 \models \mathsf{tr}_{21}\big(\mathsf{tr}_{12}(\varphi)\big)$. Consequently, $T_1 \models \varphi$ since $T_1 \models \mathsf{tr}_{21}\big(\mathsf{tr}_{12}(\varphi)\big) \leftrightarrow \varphi$; and this is what we need to show. □

Definition 2.8. Let T_1 and T_2 be two theories. We say that T_2 *is a* ***conservative extension*** *of* T_1, in symbols $T_1 \sqsubseteq T_2$, iff $\mathsf{Fm}_1 \subseteq \mathsf{Fm}_2$ and, for all $\varphi \in \mathsf{Fm}_1$,
$$T_2 \models \varphi \iff T_1 \models \varphi.$$

---

[3]Here, we need a deterministic mechanism for generating a substituted version of φ_R with no "collisions of variables". The are several ways to do this in the literature. For example, we can use Tarski's substitution of variables as follows. Let $y_0, \dots, y_m$ be chosen such that $y_i = v_{l+i}$, where l is the maximum of $0, \dots, m, i_0, \dots, i_m$. Then the appropriate substitution of φ_R in this case can be the following formula:
$$\exists v_0(v_0 = y_0 \wedge \dots \wedge \exists v_m(v_m = y_m \wedge \exists y_0(y_0 = v_{i_0} \wedge \dots \wedge \exists y_m(y_m = v_{i_m} \wedge \varphi_R)))).$$

We note that $T_1 \sqsubseteq T_2$ iff the identity translation $id : \mathsf{Fm}_1 \to \mathsf{Fm}_2$ is a faithful interpretation.

We also assume familiarity with the basic notions of set theory. For instance, what is a set, a class, a function, etc. The only difference is that in the context of this paper, there is no point to distinguish different kinds of infinities. Therefore, instead of cardinality, here we are going to speak about the ***size*** *of set* X, defined as follows:

$$||X|| \stackrel{\text{def}}{=} \begin{cases} k & X \text{ is finite and has exactly } k\text{-many elements,} \\ \infty & \text{if } X \text{ is an infinite set.} \end{cases}$$

3. Cluster networks & Step distances

Now, we introduce a general way of defining a distance on any given class X. We note that one of our targets is to define distances on the class of all theories, thus we need to work with classes which are not necessarily sets.

Definition 3.1. By a **cluster** (X, E) we mean a class X equipped with an equivalence relation E.[4]

We are interested in distances according to which some different objects are indistinguishable. Indeed, it is natural to treat equivalent theories as if they were of distance 0 from each other. As we mentioned in the introduction, there are several notions of equivalence between theories. Such equivalence thus can be represented in the above definition by the relation E.

Definition 3.2. A ***cluster network*** is a triple $(\mathrm{X}, \mathrm{E}, \mathrm{S})$, where (X, E) is a cluster and S is a symmetric relation on X.[5]

Given a cluster network $(\mathrm{X}, \mathrm{E}, \mathrm{S})$. A ***path*** *leading from* $T \in \mathrm{X}$ *to* $T' \in \mathrm{X}$ *in* $(\mathrm{X}, \mathrm{E}, \mathrm{S})$ is a finite sequence $b_1, \ldots, b_m$ of 0's and 1's such that there is a sequence $T_0, \ldots, T_m$ of members of X with $T_0 = T$, $T_m = T'$ and, for each $1 \leq i \leq m$,

$$b_i = 0 \iff T_{i-1} \mathrel{\mathrm{E}} T_i \quad \text{and} \quad b_i = 1 \iff T_{i-1} \mathrel{\mathrm{S}} T_i.$$

The ***length*** *of this path* is defined to be $\sum_{i=0}^{m} b_i$. Two objects $T, T' \in \mathrm{X}$ *are* ***connected*** *in* $(\mathrm{X}, \mathrm{E}, \mathrm{S})$ iff there is a path leading from one of them to the other in $(\mathrm{X}, \mathrm{E}, \mathrm{S})$.

Definition 3.3. Let $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{S})$ be a cluster network. The ***step distance*** *on* $\mathcal{X}$ is the function $\mathrm{d}_{\mathcal{X}} : \mathrm{X} \times \mathrm{X} \to \mathbb{N} \cup \{\infty\}$ defined as follows. For each $T, T' \in \mathrm{X}$:

- If T and T' are not connected in $(\mathrm{X}, \mathrm{E}, \mathrm{S})$, then $\mathrm{d}_{\mathcal{X}}(T, T') \stackrel{\text{def}}{=} \infty$.
- If T and T' are connected in $(\mathrm{X}, \mathrm{E}, \mathrm{S})$, then

$$\mathrm{d}_{\mathcal{X}}(T, T') \stackrel{\text{def}}{=} \min\{k : \exists \text{ a path leading from } T \text{ to } T' \text{ whose length is } k\}.$$

[4]All definitions in this section can be formulated within von Neumann–Bernays–Gödel set theory (NBG). Of course, ordered pairs of proper classes cannot be formulated even in NBG, but we do not really need ordered pairs here. Our definitions can be understood as follows: "for all classes X, E, etc. having certain properties there are classes d, etc. such that...". We use the notations $(-, -)$ only to make our statements easier to be read and our proofs easier to be followed.

[5]See footnote 4 above.

The equivalence relation E represents pairs that cannot be distinguished by the step distance, while the symmetric relation S represents the pairs of objects that are (at most) one step away from each other. The step distance then counts the minimum number of steps needed to reach an object starting from another object. We may need to say that infinitely many steps are needed, so we allow ∞ in the range of the step distance.

Example. Let X be any class, let E be the identity relation and let $\mathrm{S} = \mathrm{X} \times \mathrm{X}$. Then, $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{S})$ is a cluster network and its step distance is the following discrete distance:

$$\mathrm{d}_{\mathcal{X}}(T, T') = \begin{cases} 0 & \text{if } T = T', \\ 1 & \text{if } T \neq T'. \end{cases}$$

Theorem 3.4. Let $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{S})$ be a cluster network and let $\mathrm{d}_{\mathcal{X}} : \mathrm{X} \times \mathrm{X} \to \mathbb{N} \cup \{\infty\}$ be the step distance on $\mathcal{X}$. The following are true for each $T_1, T_2, T_3 \in \mathrm{X}$.

(a) $\mathrm{d}_{\mathcal{X}}(T_1, T_2) \geq 0$ and $\mathrm{d}_{\mathcal{X}}(T_1, T_2) = 0 \iff T_1 \mathrel{\mathrm{E}} T_2$.
(b) $\mathrm{d}_{\mathcal{X}}(T_1, T_2) = \mathrm{d}_{\mathcal{X}}(T_2, T_1)$.
(c) $\mathrm{d}_{\mathcal{X}}(T_1, T_2) \leq \mathrm{d}_{\mathcal{X}}(T_1, T_3) + \mathrm{d}_{\mathcal{X}}(T_3, T_2)$.

Proof. Let $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{S})$ and $\mathrm{d}_{\mathcal{X}}$ be as required. Let $T_1, T_2, T_3 \in \mathrm{X}$.

(a) Clearly, $\mathrm{d}_{\mathcal{X}}(T_1, T_2) \geq 0$ for any two $T_1, T_2 \in \mathrm{X}$, and $\mathrm{d}_{\mathcal{X}}(T_1, T_2) = 0$ if $T_1 \mathrel{\mathrm{E}} T_1$ because then 0 is a path from T_1 to T_2 in $\mathcal{X}$. If $\mathrm{d}_{\mathcal{X}}(T_1, T_2) = 0$, then there is a path $0, \ldots, 0$ from T_1 to T_2 in $\mathcal{X}$. So there is a sequence $T'_0, \ldots, T'_m \in \mathrm{X}$ such that $T_1 = T'_0$, $T_2 = T'_m$ and $T'_{i-1} \mathrel{\mathrm{E}} T'_i$ for each $1 \leq i \leq m$. Hence $T_1 \mathrel{\mathrm{E}} T_2$ since E is transitive.
(b) The symmetry is satisfied because E and S are symmetric relations. Hence if $b_1, \ldots, b_m$ is a path leading from T_1 to T_2 in $\mathcal{X}$, then $b_m, \ldots, b_1$ is a path leading from T_2 to T_1 in $\mathcal{X}$.
(c) The triangle inequality $\mathrm{d}_{\mathcal{X}}(T_1, T_2) \leq \mathrm{d}_{\mathcal{X}}(T_1, T_3) + \mathrm{d}_{\mathcal{X}}(T_3, T_2)$ follows from the definition because, if $b_1, \ldots, b_m$ is a path leading from T_1 to T_3 in $\mathcal{X}$ and $c_1, \ldots, c_k$ is a path leading from T_3 to T_2 in $\mathcal{X}$, then $b_1, \ldots, b_m, c_1, \ldots, c_k$ is a path leading from T_1 to T_2 in $\mathcal{X}$. □

Lemma 3.5. Let $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{S})$ and $\mathcal{X}' = (\mathrm{X}', \mathrm{E}', \mathrm{S}')$ be two cluster networks. If one of the following holds then $\mathrm{d}_{\mathcal{X}}(T_1, T_2) \geq \mathrm{d}_{\mathcal{X}'}(T_1, T_2)$ for each $T_1, T_2 \in \mathrm{X}$.

(i) $\mathrm{X} \subseteq \mathrm{X}'$, $\mathrm{E} = \mathrm{E}' \cap (\mathrm{X} \times \mathrm{X})$, and $\mathrm{S} = \mathrm{S}' \cap (\mathrm{X} \times \mathrm{X})$.
(ii) $\mathrm{X} = \mathrm{X}'$, $\mathrm{E} \subseteq \mathrm{E}'$, and $\mathrm{S} = \mathrm{S}'$.
(iii) $\mathrm{X} = \mathrm{X}'$, $\mathrm{E} = \mathrm{E}'$, and $\mathrm{S} \subseteq \mathrm{S}'$.

Proof. It follows from the fact that every path in $\mathcal{X}$ is contained in $\mathcal{X}'$ too. Since, it is just routine, we omit the details from here. □

Now, we use the above general settings to define distances between theories. Before we start, we need the following convention: Suppose that we are given two theories T and T'. We write $T \leftarrow T'$ iff there is $\varphi \in \mathsf{Fm}$ such that $T \cup \{\varphi\} \equiv T'$. We also write $T - T'$ iff either $T \leftarrow T'$ or $T' \leftarrow T$. Conventionally, we call the relation $\leftarrow$ ***axiom adding***, while the converse relation $\rightarrow$ is called ***axiom removal***. It is easy to see that the following are true for any theories T_1, T_2 and T_3.

$$T_1 \leftarrow T_2 \ \& \ T_2 \leftarrow T_3 \quad \Longrightarrow \quad T_1 \leftarrow T_3, \tag{1}$$

$$T_1 \equiv T_2 \ \& \ T_2 \leftarrow T_3 \quad \Longrightarrow \quad T_1 \leftarrow T_3, \tag{2}$$

$$T_1 \leftarrow T_2 \ \& \ T_2 \equiv T_3 \quad \Longrightarrow \quad T_1 \leftarrow T_3. \tag{3}$$

Definition 3.6. Let X be a class of some theories and consider the cluster network $(\mathrm{X}, \equiv, -)$. We call the step distance on this cluster network ***axiomatic distance*** *on* X. This step distance will be denoted by Ad_{X}.

Let X be a class of theories. We note the following. If there is a path (of finite length) between $T, T' \in \mathrm{X}$ in the cluster network $(\mathrm{X}, \equiv, -)$, then both T and T' must be formulated in the same language. In other words, if T, T' are formulated on different languages, then $\mathrm{Ad}_{\mathrm{X}}(T, T') = \infty$. This is because two theories can only be logically equivalent if they are formulated on the same language.

Example. Let X be a class of theories. Let $T, T_\perp \in \mathrm{X}$ be two theories formulated in the same language. Suppose that T is consistent while $T_\perp$ is inconsistent. Then, adding a contradiction to T ensures that $\mathrm{Ad}_{\mathrm{X}}(T, T_\perp) = 1$.

Example. Let X be a class of theories. Let $T, \emptyset \in \mathrm{X}$ be two theories formulated in the same language such that $\emptyset$ is an empty theory (i.e., empty set of formulas). Suppose that T is finitely axiomatizable, then we have either

$$\mathrm{Ad}_{\mathrm{X}}(T, \emptyset) = 1 \quad \text{or} \quad T \equiv \emptyset.$$

Thus, in the class of all theories, the axiomatic distance between any two finitely axiomatizable theories is ≤ 2.

Example. Let X be the set of all consistent theories of binary relations, let T_P be the theory of partial orders, and let T_E be the theory of equivalence relations. Then $\mathrm{Ad}_{\mathrm{X}}(T_P, T_E) = 2$. It is easy to see this because $\mathrm{Ad}_{\mathrm{X}}(T_P, T_E) \geq 2$ since none of T_p or T_E implies the other, and $\mathrm{Ad}_{\mathrm{X}}(T_P, T_E) \leq \mathrm{Ad}_{\mathrm{X}}(T_P, \emptyset) + \mathrm{Ad}_{\mathrm{X}}(\emptyset, T_E) = 2$ by Example 3 and Theorem 3.4 (c).

This gives us the intuition that the axiomatic distance is of limited interest. In most of the cases, this distance is either 0, 1, 2 or ∞. For example, in the class of all complete and consistent theories, the axiomatic distance is either 0, 1 or ∞.

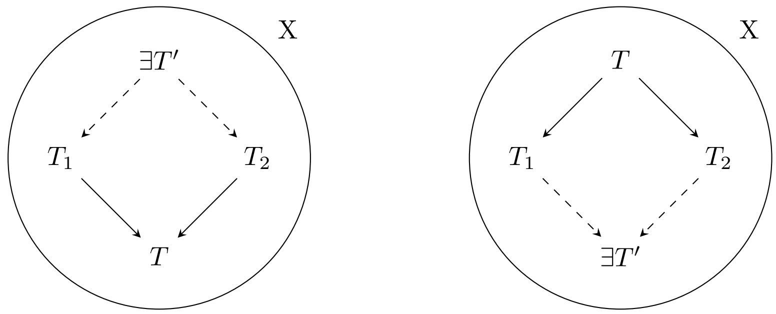
Problem 1. Let X be the class of all consistent theories. Is it true that, if the axiomatic distance between $T, T' \in \mathrm{X}$ is finite, then it must be ≤ 2?

Now, let us try to answer the above problem. We define the properties illustrated in Figure 1. Let X be a class of theories. We say that X *has the* ***theory amalgamation property*** iff for each $T \in \mathrm{X}$, if there are $T_1, T_2 \in \mathrm{X}$ such that $T_1 \to T \leftarrow T_2$, then there is $T' \in \mathrm{X}$ such that $T_1 \leftarrow T' \to T_2$. Analogously, we say that X has the ***theory co-amalgamation property*** iff for each $T \in \mathrm{X}$, if there are $T_1, T_2 \in \mathrm{X}$ such that $T_1 \leftarrow T \to T_2$, then there is $T' \in \mathrm{X}$ such that $T_1 \to T' \leftarrow T_2$.

Theorem 3.7. Let X a class of theories having the theory amalgamation property or the theory co-amalgamation property. Then for all $T, T' \in \mathrm{X}$, we have the following

$$\mathrm{Ad}_{\mathrm{X}}(T, T') = \begin{cases} 0 & \text{if } T \equiv T', \\ 1 & \text{if } T' \text{ or } T \text{ is finitely axiomatizable over the other,} \\ \infty & \text{if } T \text{ and } T' \text{ are not connected in } (\mathrm{X}, \equiv, -), \\ 2 & \text{otherwise.} \end{cases} \tag{4}$$

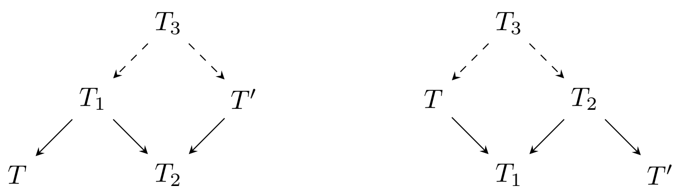
Proof. Let us first assume that X has the theory amalgamation property. Suppose that $T, T' \in \mathrm{X}$ are connected via a path of length 3 in the cluster network $(\mathrm{X}, \equiv, -)$. By (2) and (3), we can find



(a) amalgamation property (b) co-amalgamation property

Figure 1. Theory amalgamation properties

$T_1, T_2 \in \mathrm{X}$ such that $T - T_1 - T_2 - T'$. Note that T, T_1, T_2, T' have the same language $\mathcal{L}$ and the same set of formulas Fm. We first show that T and T' are connected by a path of length 2. If at least two consecutive $-$ in the path $T - T_1 - T_2 - T'$ are in the same direction, e.g., $T \leftarrow T_1 \rightarrow T_2 \rightarrow T'$, then we are done by (1). So, we may assume that we have one the cases of Figure 2:

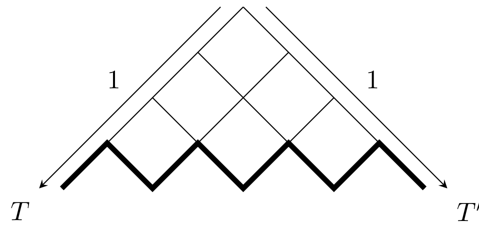


(a) Case 1 (b) Case 2

Figure 2. 3-paths can be replaced by 2-paths

(a) Suppose that we are in the first case $T \leftarrow T_1 \rightarrow T_2 \leftarrow T'$. Then by the theory amalgamation property, there is $T_3 \in \mathrm{X}$ such that $T_1 \leftarrow T_3 \rightarrow T'$. Hence, by (1), we have $T \leftarrow T_3 \rightarrow T'$ which means T and T' are connected by a path of length 2.
(b) Suppose that we are in the second case $T \rightarrow T_1 \leftarrow T_2 \rightarrow T'$. Then by the theory amalgamation property, there is $T_3 \in \mathrm{X}$ such that $T \leftarrow T_3 \rightarrow T_2$. Hence, by (1), we have $T \leftarrow T_3 \rightarrow T'$ which means T and T' are connected by a path of length 2.

Therefore, any path of length 3 can be replaced by a path of length 2. Now, we can prove the theorem. The first three cases of (4) are obvious, we need to show that otherwise the axiomatic distance is 2. We use induction on on the length l of path. If the path is of length $l = 3$, then we are done by the above discussion. Suppose that we have already proven that every path not longer than $l \geq 3$ can be replaced by a path of length 2. Let path $T - T_1 - \cdots - T_n - T'$ be a path of length $l + 1$. By induction hypothesis, path $T - T_1 - \cdots - T_n$ can be replaced by a path of length 2. Hence, path $T - T_1 - \cdots - T_n - T'$ can be replaced by a path of length 3, which can be replaced by a path of length 2 by the induction hypothesis.



The proof of the case when X has the theory co-amalgamation property is completely analogous, but going downwards instead of upwards. □

Even though Theorem 3.7 shows the simplicity of axiomatic distance in plenty of cases, it leaves open the important case when X is the class of all consistent theories (recall Problem 1).

Proposition 3.8. Let X be the class of all consistent theories formulated over a fixed language $\mathcal{L}$. Then, the class X does not have the theory amalgamation property. Moreover, if $\mathcal{L}$ contains at least a binary relation symbol, then X does not have the theory co-amalgamation property either.

Proof. Let $\mathcal{L}$ be a language and let X be the class of all consistent theories of language $\mathcal{L}$. Let $T = \emptyset$, $T_1 = \{\exists x(x = x)\}$ and $T_2 = \{\forall x(x \neq x)\}$ be three theories in X. Then clearly, $T_1 \to T \leftarrow T_2$, but there is no consistent theory $T' \in$ X for which $T_1 \leftarrow T' \to T_2$. Hence X does not have the theory amalgamation property.

Now assume that $\mathcal{L}$ contains at least one binary relation symbol R. Let $T_1 = \{\neg\Psi^{(2n)} : n \in \mathbb{N}\}$ and let $T_2 = \{\neg\Psi^{(2n+1)} : n \in \mathbb{N}\}$ be two theories in X, where $\Psi^{(n)}$ is a formula in language $\mathcal{L}$ saying that there are exactly n-many objects, see page 14 below for precise formulas capturing this. Let φ be a formula in language $\mathcal{L}$ requiring that there are infinitely many objects. Using the relation symbol R it is easy to write up such a formula by requiring that R is irreflexive, transitive and serial. Then clearly, $T_1 \cup \{\varphi\} \equiv T_2 \cup \{\varphi\} \equiv \{\varphi\}$. Hence $T = \{\varphi\}$ is a theory for which $T_1 \leftarrow T \to T_2$. However, there is no theory T' for which $T_1 \to T' \leftarrow T_2$ because (1) $\mathsf{Cn}(T_1) \cap \mathsf{Cn}(T_2) \equiv \emptyset$ since every model $\mathfrak{M}$ is either a model for T_1 or a model for T_2, and (2) T_1 and T_2 are not finitely axiomatizable because their complements are not closed under ultraproducts, see [**?**, Corollary 9.5.10, p.454]. Consequently, X does not have the theory co-amalgamation property. □

It is worth mentioning that if we replace every occurrence of the logical equivalence $\equiv$ by the definitional equivalence $\rightleftarrows$ then a similar result can be proven in the same way.

4. Conceptual distance

We have shown that axiomatic distance can be simple in many interesting cases. So instead of logical equivalence, here we consider definitional equivalence. We define conceptual distance between theories, and we show that it is more interesting than axiomatic distance, in a sense.

Definition 4.1. Let T and T' be theories. We say that T' *is a* ***one-concept-extension*** *of* T and we write $T \rightsquigarrow T'$ iff $\mathcal{L}' = \mathcal{L} \cup \{R\}$, for some relation symbol R, and $T \sqsubseteq T'$. We also write $T \sim T'$ iff $T \rightsquigarrow T'$ or $T' \rightsquigarrow T$, and in this case we say that T *and* T' *are* ***separated by at most one concept***.

Again, we have not define a separate notion of concept removal, instead we understand concept removal to be the converse of concept adding. Later, in Section 6, we are going to introduce another notion for concept removal, and a corresponding distance notion that we call bi-directed conceptual distance.

Definition 4.2. Let X be a class of theories. The step distance induced by the cluster network $(\mathrm{X}, \rightleftarrows, \sim)$ is called ***conceptual distance*** *on* X and is denoted by Cd_{X}. In the case when X is the class of all theories, we omit the subscript and we write Cd.

Intuitively, by a concept we mean a definition, no matter how many different ways one can write it equivalently. It is apparent that an inconsistent theory is of an infinite conceptual distance from any consistent theory, because $\rightleftarrows$ and $\rightsquigarrow$ cannot make a consistent theory inconsistent and also cannot make an inconsistent theory consistent.

Definition 4.3. Let T be a theory and let $k \in \mathbb{N} \cup \{\infty\}$. We say that the ***conceptual size*** *of* T *is* k and we write $\mathsf{Cz}(T) = k$ iff there is a maximal set $X \subseteq \mathsf{Fm}$ such that $||X|| = k$ and $T \not\models \varphi \leftrightarrow \psi$ for any $\varphi, \psi \in X$.[6] We call T ***conceptually finite*** iff its conceptual size $\mathsf{Cz}(T)$ is finite.

Let us note that a theory that has a model of at least two different elements cannot be conceptually finite (because we have infinitely many variables). We call a theory T ***propositional theory*** iff $T \models \varphi \leftrightarrow \exists x\, \varphi$ for each $\varphi \in \mathsf{Fm}$. It is worthy of note that, if a theory T is conceptually finite, then it is propositional. However, this is not true if we consider other variants of FOL, e.g., finite-variable fragments.

Lemma 4.4. Let T, T' be two arbitrary theories. If there is a faithful interpretation from T to T', then $\mathsf{Cz}(T) \leq \mathsf{Cz}(T')$. Consequently,

$$T \rightleftarrows T' \implies \mathsf{Cz}(T) = \mathsf{Cz}(T').$$

Proof. Let T and T' be two arbitrary theories, and suppose that tr is a faithful interpretation of T into T'. By definition, there is a maximal set $X \subseteq \mathsf{Fm}$ of size $\mathsf{Cz}(T)$ such that $T \not\models \varphi \leftrightarrow \psi$ for any $\varphi, \psi \in X$. Let $X' \subseteq \mathsf{Fm}'$ be the following set $X' = \{\mathsf{tr}(\varphi) : \varphi \in X\}$. Since tr is a faithful interpretation, $T' \not\models \mathsf{tr}(\varphi \leftrightarrow \psi)$. Hence $T' \not\models \mathsf{tr}(\varphi) \leftrightarrow \mathsf{tr}(\psi)$ since tr is an interpretation. Consequently, $\mathsf{Cz}(T) \leq \mathsf{Cz}(T')$. If $T \rightleftarrows T'$, then by Proposition 2.7 there are faithful interpretations between T and T' in both directions. Hence $\mathsf{Cz}(T) \leq \mathsf{Cz}(T') \leq \mathsf{Cz}(T)$. Therefore, $\mathsf{Cz}(T) = \mathsf{Cz}(T')$ and the desired follows. □

Now, as a preparation of showing that conceptual distance can be arbitrary, we compute the conceptual size of some theories. To do this, we borrow some ideas from the theory of Boolean algebras.

Definition 4.5. Suppose that T is a theory. An ***atomic concept*** *of theory* T is a formula $\alpha \in \mathsf{Fm}$ with the following conditions:

(a) α is satisfiable in theory T.
(b) For each formula $\varphi \in \mathsf{Fm}$, exactly one of these formulas $\alpha \wedge \varphi$ and $\alpha \wedge \neg\varphi$ is a contradiction to T.

[6] It might be more convenient in some cases to define the conceptual size as the cardinality of different concepts of theory T.

Let T be a theory, and let $\alpha, \varphi \in Fm$. We note that if $\alpha \wedge \neg\varphi$ is a contradiction to T, then $T \models \alpha \to \varphi$. So, in this case, we will say that α *is below* φ *in* T.

Lemma 4.6. Let T be a consistent and conceptually finite theory. Then there is a ***complete set*** *of atomic concepts of theory* T. That means, there is a finite set $\mathsf{Atoms} = \{\alpha_1, \ldots, \alpha_m\}$ of atomic concepts of T (for some $m \in \mathbb{N}$) such that:

(a) For every different $\alpha_i, \alpha_j \in \mathsf{Atoms}$, the conjunction $\alpha_i \wedge \alpha_j$ is a contradiction to theory T.
(b) For each formula $\varphi \in \mathsf{Fm}$ there is $\Sigma \subseteq \mathsf{Atoms}$ with $T \models \varphi \leftrightarrow \bigvee \Sigma$.

Proof. Let T be a consistent and conceptually finite theory. We first prove the following.

(*) For every formula $\varphi \in \mathsf{Fm}$ which is satisfiable in theory T, there is an atomic concept α of T which is below φ in T.

Let $\varphi \in \mathsf{Fm}$ be a satisfiable formula in T. Suppose towards a contradiction that such atomic concept (mentioned in (*)) does not exist. Thus, φ itself is not an atomic concept and so there is $\psi_1 \in \mathsf{Fm}$ such that $\varphi \wedge \psi_1$ is satisfiable and $T \not\models \varphi \leftrightarrow \varphi \wedge \psi_1$ (because $\varphi \wedge \neg\psi_1$ is also satisfiable). Again, $\varphi \wedge \psi_1$ is not an atomic concept. Hence, we can find $\psi_2 \in \mathsf{Fm}$ such that $\varphi \wedge \psi_1 \wedge \psi_2$ is satisfiable, $T \not\models \varphi \wedge \psi_1 \leftrightarrow \varphi \wedge \psi_1 \wedge \psi_2$ and $T \not\models \varphi \leftrightarrow \varphi \wedge \psi_1 \wedge \psi_2$. If we continue in this procedure, we will get infinitely many formulas $\psi_1, \psi_2, \ldots \in \mathsf{Fm}$ such that each of the formulas

$$\varphi, \quad \varphi \wedge \psi_1, \quad \varphi \wedge \psi_1 \wedge \psi_2, \quad \ldots$$

is satisfiable and none of them is equivalent to another one in theory T. This makes a contradiction with the conceptual finiteness of theory T. We have showed that (*) is true. Now, suppose that $\mathsf{Atoms} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is a maximal set of atomic concepts whose elements are different (mutually non-equivalent) in theory T. We can suppose that Atoms is finite because T is conceptually finite. We claim that Atoms is the desired complete set of atomic concepts. Let $\alpha_i, \alpha_j \in \mathsf{Atoms}$ be such that $\alpha_i \neq \alpha_j$. Suppose that $\alpha_i \wedge \alpha_j$ is satisfiable. Then, by Definition 4.5, both formulas $\alpha_i \wedge \neg\alpha_j$ and $\neg\alpha_i \wedge \alpha_j$ are contradictions. Therefore,

$$T \models \alpha_i \to \alpha_j \quad \text{and} \quad T \models \alpha_j \to \alpha_i,$$

which contradicts the choice of the set Atoms. It remains to prove the following.

(**) For each formula $\varphi \in \mathsf{Fm}$, there is a subset $\Sigma \subseteq \mathsf{Atoms}$ such that

$$T \models \varphi \leftrightarrow \bigvee \Sigma.$$

Let $\varphi \in \mathsf{Fm}$, and let $\Sigma \subseteq \mathsf{Atoms}$ be the set of all atomic concepts in Atoms that are below φ in T. The direction $T \models \varphi \leftarrow \bigvee \Sigma$ is clear because each element of Σ is below φ. Suppose towards a contradiction that the other direction is not true. Then, $\varphi \wedge \bigwedge\{\neg\alpha : \alpha \in \Sigma\}$ is satisfiable in T, and by (*), there is must be an atomic concept β below $\varphi \wedge \bigwedge\{\neg\alpha : \alpha \in \Sigma\}$ in T. Hence, $T \not\models \beta \leftrightarrow \alpha$ for each $\alpha \in \mathsf{Atoms}$. This contradicts the maximality of Atoms. Thus, $T \models \varphi \leftrightarrow \bigvee \Sigma$. Therefore, (**) holds and the statement of the lemma follows. □

Corollary 4.7. Let T be a consistent and conceptually finite theory. Then, there is a finite number $m \in \mathbb{N}$ such that $\mathsf{Cz}(T) = 2^m$.

Proof. Let T be a consistent and conceptually finite theory. Suppose that Atoms is a complete set of atomic concepts for theory T (given by Lemma 4.6). Let $\Sigma, \Sigma' \subseteq \mathsf{Atoms}$ be such that $\Sigma \neq \Sigma'$. We need to prove that $T \not\models \bigvee \Sigma \leftrightarrow \bigvee \Sigma'$. Without loose of generality, we may assume that there is an atomic concept $\alpha \in \mathsf{Atoms}$ such that $\alpha \in \Sigma$ and $\alpha \notin \Sigma'$. Thus, by the completeness of the set Atoms, formula $\bigvee \Sigma \wedge \alpha$ is satisfiable in T while formula $\bigvee \Sigma' \wedge \alpha$ is a contradiction to T. Therefore, it follows that $T \not\models \bigvee \Sigma \leftrightarrow \bigvee \Sigma'$ as desired. □

Lemma 4.8. Let T, T' be two propositional theories. Then,

$$T \rightsquigarrow T' \implies \mathsf{Cz}(T) \leq \mathsf{Cz}(T') \leq (\mathsf{Cz}(T))^2.$$

Proof. Let T, T' be two propositional theories and suppose that $T \rightsquigarrow T'$. We may also assume that both T and T' are consistent. If $\mathsf{Cz}(T)$ is infinite, then we are done. So, suppose that T is conceptually finite. Let the complete set of atomic concepts given by Lemma 4.6 be $\mathsf{Atoms} = \{\alpha_1, \ldots, \alpha_m\}$. Suppose that $\mathcal{L}' = \mathcal{L} \cup \{R\}$; such R exists by the fact that $T \rightsquigarrow T'$. Let

$$A = \{\alpha_1 \wedge \bar{R}, \ldots, \alpha_m \wedge \bar{R}, \alpha_1 \wedge \neg\bar{R}, \ldots, \alpha_m \wedge \neg\bar{R}\}.$$

Here, $\bar{R} \stackrel{\text{def}}{=} R(v_0, \ldots, v_0)$. By a simple induction on the complexity of the formulas in Fm', one can easily prove that

$$(\forall \psi \in Fm') \ \ (\exists \Sigma \subseteq A) \quad T' \models \psi \leftrightarrow \bigvee \Sigma. \tag{5}$$

To proceed with the induction, all what we need are the following two facts: (1) A forms a partition of a (any) theorem of T'. In other words, $\bigvee A$ is a theorem of T' and, for each $\beta_1, \beta_2 \in A$, if $\beta_1 \neq \beta_2$ then $\beta_1 \wedge \beta_2$ is a contradiction to T'. (2) Theory T' is propositional. We omit the details as it is just a routine.

Let us note that A has exactly $2m$ elements (as a set). So, by (5), it follows that T' can have at most $2^{2m} = (2^m)^2 = (\mathsf{Cz}(T))^2$ many different concepts. We note that the equality $\mathsf{Cz}(T') = (\mathsf{Cz}(T))^2$ would happen iff none of the elements in A is contradicting T, while $\mathsf{Cz}(T') = \mathsf{Cz}(T)$ would happen iff $\bar{R}$ is a theorem of T or a contradiction to T. □

Let $\mathcal{L}_\infty$ be the language that consists of infinitely many unary relation symbols (describing infinitely many different concepts, say colors)

$$c_1, c_2, \ldots \tag{6}$$

For each $n \in \mathbb{N}$, we let $\mathcal{L}_n \subseteq \mathcal{L}_\infty$ be the language consists of the first n-many relation symbols from the list in (6). For each $k \in \mathbb{N} \cup \{\infty\}$, we let $P_k = \{x = y\}$ be a theory on language $\mathcal{L}_k$. Note that all the theories $P_\infty, P_0, P_1, \ldots$ are in fact propositional theories.

Proposition 4.9. Let $k \in \mathbb{N} \cup \{\infty\}$. Then $\mathrm{Cd}(P_0, P_k) = k$.

Proof. Since $P_i \rightsquigarrow P_{i+1}$ for all $i \in \mathbb{N}$, it immediately follows that $\mathrm{Cd}(P_0, P_k) \leq k$. To show that $\mathrm{Cd}(P_0, P_k) \geq k$, we should show that in any path going from P_0 to P_k (in the cluster network of all theories) there is at least k many one-concept-extension steps. We first note that all the theories in such a path must be propositional. Indeed, P_k is propositional, a conservative extension of a propositional theory is propositional theory, too, and the same is true for the definitional equivalence (one can also see Proposition 4.11 below for an alternative proof). Moreover, one can check that

$$\{c_1 \wedge c_2 \wedge \ldots \wedge c_k, \neg c_1 \wedge c_2 \wedge \ldots \wedge c_k, \ldots, \neg c_1 \wedge \neg c_2 \wedge \ldots \wedge \neg c_k\}$$

is a complete set of atomic concepts of theory P_k. Hence, $\mathsf{Cz}(P_k) = 2^{2^k}$. By Lemma 4.4 and Lemma 4.8, it is apparent that at least k one-concept-extension steps are needed to increase the conceptual size $\mathsf{Cz}(P_0) = 2$ to $\mathsf{Cz}(P_k) = 2^{2^k}$. □

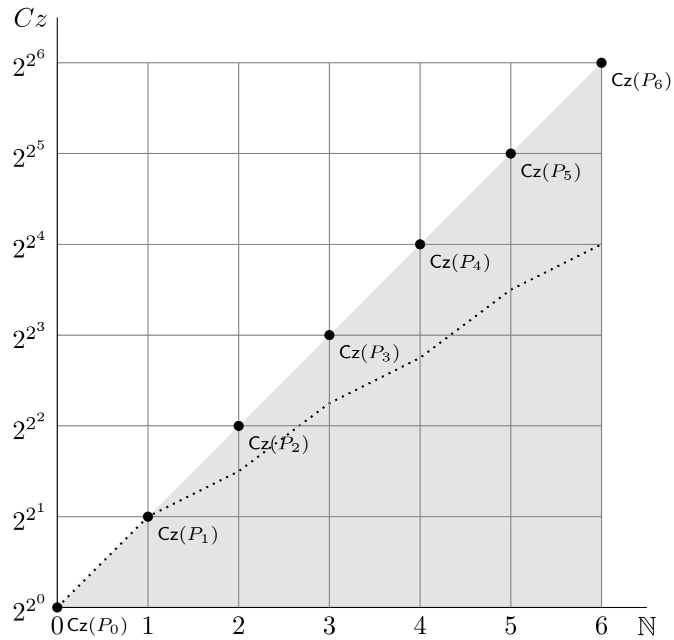


FIGURE 3. Illustration for the proof of Proposition 4.9

Remark 1. The same ideas used in the above proof can be used to prove the following general observation: For every $n, m \in \mathbb{N}$,

$$\mathrm{Cd}(P_n, P_m) = |n - m| \quad \text{and} \quad \mathrm{Cd}(P_n, P_\infty) = \infty.$$

Theorem 4.10. For each $k \in \mathbb{N} \cup \{\infty\}$, there are theories T and T' of conceptual distance k, i.e., $\mathrm{Cd}(T, T') = k$.

Proof. Follows immediately from Proposition 4.9 □

Now, we introduce an interesting implementation of conceptual distance in model theory. For every $n \in \mathbb{N}$, consider the following formula saying that there are exactly n many distinct objects

$$\Psi^{(n)} \stackrel{\text{def}}{=} \exists x_1 x_2 \dots x_n \left(\bigwedge_{i \neq j} x_i \neq x_j \wedge \forall y \left[\bigvee_i y = x_i \right] \right).$$

Note that a theory T is propositional iff $T \models \Psi^{(1)}$.

Proposition 4.11. Let T and T' be two theories such that $T \models \Psi^{(n)}$. Then

$$\mathrm{Cd}(T, T') < \infty \implies T' \models \Psi^{(n)}.$$

In particular, if theory T is a propositional theory and $\mathrm{Cd}(T, T') < \infty$, then T' is also propositional.

Proof. The statement follows from the following simple observation: if $T \sqsubseteq T'$ or $T \rightleftarrows T'$, then $T \models \Psi^{(n)}$ iff $T' \models \Psi^{(n)}$. □

Definition 4.12. Let α be a cardinality. The ***spectrum*** *of theory* T, in symbols $I(T,\alpha)$, is the number (up to isomorphism) of its models of cardinality α.

Proposition 4.13. Let α be a cardinality and let T and T' be two theories formulated on a countable language. If $\mathrm{Cd}(T,T') < \infty$, then T has a model of cardinality α iff T' has a model of cardinality α, i.e.,

$$\mathrm{Cd}(T,T') < \infty \implies \big[I(T,\alpha) \neq 0 \iff I(T',\alpha) \neq 0\big].$$

Proof. Let α, T and T' be as required. It is enough to show that if T has (does not have) a model of cardinality α and $T' \rightleftarrows T$, $T' \sqsubseteq T$ or $T \sqsubseteq T'$, then T' also has (does not have) a model of cardinality α.

Suppose that α is a finite cardinality, then T does not have a model over α iff $T \models \neg\Psi^{(\alpha)}$. If $T' \sqsubseteq T$ or $T \sqsubseteq T'$, then $T' \models \neg\Psi^{(\alpha)}$ iff $T \models \neg\Psi^{(\alpha)}$ because $\Psi^{(\alpha)}$ is formulated in the common language. If $T' \rightleftarrows T$, then

$$T' \models \neg\Psi^{(\alpha)} \iff T \models \neg\Psi^{(\alpha)}$$

because $\mathsf{tr}(\neg\Psi^{(\alpha)}) = \neg\Psi^{(\alpha)}$ for any translation tr. Hence, T' has a model of cardinality α iff T has a model of cardinality α.

If α is an infinite cardinality, then by Lövenheim–Skolem Theorem, T and T' have a model of cardinality α iff they have an infinite model. T has an infinite model iff $T \models \exists x_1 x_2 \dots x_n \left(\bigwedge_{i\neq j} x_i \neq x_j\right)$ for all $n \in \mathbb{N}$. For the same reasons as above, if $T' \rightleftarrows T$, $T' \sqsubseteq T$ or $T \sqsubseteq T'$, then $T' \models \exists x_1 x_2 \dots x_n \left(\bigwedge_{i\neq j} x_i \neq x_j\right)$ iff $T \models \exists x_1 x_2 \dots x_n \left(\bigwedge_{i\neq j} x_i \neq x_j\right)$. Hence theory T has an infinite model and thus a model of cardinality α iff theory T' has such a model. □

Problem 2. Can the conceptual distance Cd be characterized, at least for some particular cases, in terms of spectrum of theories?

Corollary 4.14. The conceptual distance between the theories of any two finite models of different cardinalities is infinite. More precisely, if $\mathfrak{A}$ and $\mathfrak{B}$ are two finite models of different cardinality, then $\mathrm{Cd}\big(\mathsf{Th}(\mathfrak{A}), \mathsf{Th}(\mathfrak{B})\big) = \infty$.

For instance, given two cyclic groups $\langle k_1 \rangle$ and $\langle k_2 \rangle$ of orders 5 and 7, respectively, the conceptual distance between the theories of these groups is ∞. This might seem strange; these theories are talking about similar structures. But if we look carefully at the statement of the above corollary, we will find that it talks about theories of structures, not the structures themselves. In other words, the conceptual distance between the theories of $\langle k_1 \rangle$ and $\langle k_2 \rangle$ cannot be granted as a distance between these two groups as algebraic structures. This conceptual distance can be rather considered as a distance between the Lindenbaum-Tarski algebras of the theories of these groups, which are of course of different nature than the groups themselves. This is planned to be discussed in details in a forthcoming algebraic oriented paper.

Corollary 4.15. There are infinitely many theories that are conceptually infinitely far from each other.

Problem 3. Let X be the class of all complete and consistent theories, and let $T_1, T_2 \in \mathrm{X}$. Is it always true that

$$\mathrm{Cd}(T_1,T_2) = \mathrm{Cd}_{\mathrm{X}}(T_1,T_2)?$$

5. Conceptual distance in physics

Each physical theory is established based on some preliminary decisions. These decisions are suggested by the accumulation and the assimilation of new knowledge. The methods used to improve the physical theories are intuitively conceived and applied in a fruitful way, but many obvious ambiguities have appeared. To clarify these ambiguities, it was critical to introduce the *logical foundation of physical theories*.

Even today the logic based axiomatic foundations of physical theories is intensively investigated by several research groups. For example, the Andréka–Németi school axiomatize and investigate special and general relativity theories within first order logic, see, e.g., [Andréka et al., 2002], [Andréka et al., 2004] and [Andréka et al., 2012]. For similar approaches related to other physical theories, see, e.g., [Baltag and Smets, 2005], [Krause and Arenhart, 2017].

Following Andréka–Németi school's traditions, two theories $\mathsf{ClassicalKin}$ and $\mathsf{SpecRel}$ are formulated in first order logic to capture the intrinsic structures of classical and relativistic kinematics. For the precise definitions of these theories, one can see [Lefever and Székely, 2018a, p.67 and p. 69]. In this section, we will investigate the conceptual distance between these two theories.

In [Lefever, 2017] and [Lefever and Székely, 2018a], it was shown that these two theories can be turned definitionally equivalent by the following two concept manipulating steps:

(1) adding the concept of an observer "being stationary" to the theory of relativistic kinematics $\mathsf{SpecRel}$ and
(2) removing the concept of observers "not moving slower than light" from the theory of classical kinematics $\mathsf{ClassicalKin}$.

Then, it was shown that even if observers "not moving slower than light" are removed from $\mathsf{ClassicalKin}$ the resulting theory remains definitionally equivalent to $\mathsf{ClassicalKin}$ and hence adding only the concept of "being stationary" to $\mathsf{SpecRel}$ is enough to make the two theories equivalent. Thus, it follows that the conceptual distance between relativistic and classical kinematics is 1.

Theorem 5.1. Classical and relativistic kinematics are distinguished from each other by only one concept, namely the existence of some distinguished observers captured by formula (7) below, i.e., $\mathrm{Cd}(\mathsf{ClassicalKin}, \mathsf{SpecRel}) = 1$.

Proof. The key to this result is the surprising theorem stating that the only concept which needs to be added to $\mathsf{SpecRel}$ to make it definitional equivalent to $\mathsf{ClassicalKin}$ is a concept distinguishing a set of observers that are "being at absolute rest" as proven in [Lefever, 2017, p.72] and [Lefever and Székely, 2018a, p.110]. Let E be the basic unary relation corresponding to this basic concept. Axiom $\mathsf{AxPrimitiveEther}$, see [Lefever, 2017, p.46] and [Lefever and Székely, 2018a, p.87], defines E as:

$$\exists e\big[\mathsf{IOb}(e) \land \forall k\big(E(k) \leftrightarrow [\mathsf{IOb}(k) \land \varphi(e,k)]\big)\big], \tag{7}$$

where IOb is a basic unary relation for inertial observers and $\varphi(e,k)$ is a formula in the language of $\mathsf{SpecRel}$ capturing that observers e and k are stationary with respect to each other. Let

$$\mathsf{SpecRel}^{\mathsf{E}} = \mathsf{SpecRel} \cup \{\mathsf{AxPrimitiveEther}\}.$$

First we are going to prove that $\mathsf{SpecRel} \rightsquigarrow \mathsf{SpecRel}^{\mathsf{E}}$. To do so, it is enough to show that $\mathsf{SpecRel} \sqsubseteq \mathsf{SpecRel}^{\mathsf{E}}$, i.e., $\mathsf{SpecRel}^{\mathsf{E}}$ is a conservative extension of $\mathsf{SpecRel}$ because the languages of these theories differ only in the unary relation symbol E. So it is enough to show that for all formula ρ of the language of $\mathsf{SpecRel}$,

$$\mathsf{SpecRel} \models \rho \iff \mathsf{SpecRel}^{\mathsf{E}} \models \rho.$$

Let ρ be an arbitrary formula of the language of $\mathsf{SpecRel}$. Since $\mathsf{SpecRel} \subseteq \mathsf{SpecRel}^{\mathsf{E}}$, $\mathsf{SpecRel} \models \rho$ implies $\mathsf{SpecRel}^{\mathsf{E}} \models \rho$. We prove the converse direction by proving that, if $\mathsf{SpecRel} \not\models \rho$, then $\mathsf{SpecRel}^{E} \not\models \rho$. Let $\mathfrak{M}$ be a model of $\mathsf{SpecRel}$. Since $\mathsf{SpecRel} \models \exists x \mathsf{IOb}(x)$, there exists an $a \in \mathsf{IOb}^{\mathfrak{M}}$. Let us fixed such element a of $\mathsf{IOb}^{\mathfrak{M}}$ and let extension $\mathfrak{M}'$ of $\mathfrak{M}$ be defined by adding the following relation to $\mathfrak{M}$:

$$E^{\mathfrak{M}'} = \left\{ b \in \mathsf{IOb}^{\mathfrak{M}} : (a,b) \in \varphi^{\mathfrak{M}} \right\},$$

where $\varphi^{\mathfrak{M}}$ is the binary relation defined by formula $\varphi(x,y)$ in model $\mathfrak{M}$. By construction, $\mathfrak{M}'$ is a model of $\mathsf{SpecRel}^{\mathsf{E}}$. Therefore, if $\mathfrak{M} \models \neg\rho$, then $\mathfrak{M}' \models \neg\rho$. Consequently, $\mathsf{SpecRel} \not\models \rho$ implies $\mathsf{SpecRel}^{\mathsf{E}} \not\models \rho$, which is what we wanted to prove. This completes the proof of $\mathsf{SpecRel} \rightsquigarrow \mathsf{SpecRel}^{\mathsf{E}}$, and hence

$$\mathrm{Cd}(\mathsf{SpecRel}, \mathsf{SpecRel}^{\mathsf{E}}) \leq 1.$$

By Corollary 9 in [Lefever, 2017, p.72] and [Lefever and Székely, 2018a, p.110], $\mathsf{SpecRel}^{\mathsf{E}}$ is definitionally equivalent to $\mathsf{ClassicalKin}$. Hence,

$$\mathrm{Cd}(\mathsf{SpecRel}^{\mathsf{E}}, \mathsf{ClassicalKin}) = 0.$$

Therefore,

$$\mathrm{Cd}(\mathsf{SpecRel}, \mathsf{ClassicalKin}) \leq 1.$$

However, $\mathrm{Cd}(\mathsf{SpecRel}, \mathsf{ClassicalKin})$ cannot be 0 since $\mathsf{SpecRel}$ and $\mathsf{ClassicalKin}$ are not definitionally equivalent, see Theorem 5 in [Lefever, 2017] or [Lefever and Székely, 2018a]. Consequently,

$$\mathrm{Cd}(\mathsf{SpecRel}, \mathsf{ClassicalKin}) = 1$$

and this is what we wanted to prove. □

Remark 2. From the ideas used in the proof of Theorem 5.1, the following general statement can be synthesized. Let T be a consistent theory, and let $\mathcal{L}'$ be the new language we can get after we add a new n-ary relation symbol P to $\mathcal{L}$. Let

$$T' = T \cup \left\{ \exists \bar{x} \left[\delta(\bar{x}) \wedge \forall \bar{y} \big(P(\bar{y}) \leftrightarrow [\delta(\bar{y}) \wedge \gamma(\bar{x},\bar{y})] \big) \right] \right\},$$

for some $\delta, \varphi \in \mathsf{Fm}$, such that $T \models \exists \bar{x} \delta(\bar{x})$. Then $T \rightsquigarrow T'$, and hence $\mathrm{Cd}(T, T') \leq 1$.

There are several ways how one can capture the structures of relativistic and classical kinematics in mathematical logic. Let us now introduce another way to capture these theories. Let $\mathbb{R}$ be the set of all real numbers. Let $\mathsf{Ph} \subseteq \mathbb{R}^4 \times \mathbb{R}^4$ be such that $(\bar{x}, \bar{y}) \in \mathsf{Ph}$ iff coordinate points $\bar{x}$ and $\bar{y}$ can be connected by a light signal, i.e., if $(x_1 - y_1)^2 - (x_2 - y_2)^2 - (x_3 - y_3)^2 - (x_4 - y_4)^2 = 0$. Let $\mathsf{S} \subseteq \mathbb{R}^4 \times \mathbb{R}^4$ be the simultaneity relation, i.e., $(\bar{x}, \bar{y}) \in S$ iff $x_1 = y_1$. Consider the models $\mathfrak{R} = \langle \mathbb{R}^4, \mathsf{Ph} \rangle$ and $\mathfrak{N} = \langle \mathbb{R}^4, \mathsf{S}, \mathsf{Ph} \rangle$, these models capture the structure of special relativity and classical kinematics, respectively.

Let $T_n = \mathsf{Th}(\mathfrak{N})$ and $T_r = \mathsf{Th}(\mathfrak{R})$. Note that T_r is in fact a conservative extension of T_n and the conceptual distance between them is 1, i.e., $\mathrm{Cd}(T_n, T_r) = 1$.

Problem 4 (Hajnal Andréka)**.** Let X be the class of all extensions of T_r that are definitionally equivalent to a subtheory of T_n. Is the following true: For all $T \in X$,

$$\mathrm{Cd}(T_r, T) + \mathrm{Cd}(T, T_n) = 1?$$

If the answer to the question in the above problem is yes, then no matter which classical (i.e., T_n-definable, but not T_r-definable) concept we add to special relativity (T_r) we will get classical kinematics (T_n). That would be an interesting insight understanding the connection between classical and relativistic concepts.

The investigation in this section opens so many questions: For any two concrete theories of physics, what is the conceptual distance between them? By Theorem 5.1, relativistic and classical kinematics are of conceptual distance one. However, the question "what is the distance between relativistic and classical dynamics?" remains open. Another natural related open problem is the following.

Problem 5 (Jean Paul Van Bendegem)**.** What is the conceptual distance between classical and statistical thermodynamics?

Of course, any answer to the above problems depends on the chosen axiomatizable theories capturing the physical theories in question. For an axiomatic approach of these thermodynamics theories, one can see, e.g., [Carathéodory, 1909], [Cooper, 1967] and [Lieb and Yngvason, 2000].

6. Ideas for other distances

Interpreting a theory into another one is a fundamental concept in logic. In the following definition, we define some distance that uses interpretations as the minimal step between theories.

Definition 6.1. For any two theories T_1 and T_2, we write $T_1 \mathbb{I} T_2$ iff one of these theories can be interpreted into the other one. Let X be an arbitrary class of theories. Note that $(\mathrm{X}, \equiv, \mathbb{I})$ is a cluster network; we call its step distance (defined in section 3) ***interpretation distance*** *on* X.

Another natural idea one may have for defining a distance between theories is using a step that collapses two concepts into one, i.e., using the symmetric closure of following relation for the minimal single steps.

Definition 6.2. Let T, T' be two theories. We say that T' *is the resultant of* T *after* ***collapsing two concepts*** iff $T' \equiv T \cup \{\varphi \leftrightarrow \psi\}$, for some $\varphi, \psi \in \mathsf{Fm}$.

Example. The theory of abelian groups is the resultant of the theory of all groups after collapsing the concepts $a \cdot b = c$ and $b \cdot a = c$.

One can easily see that collapsing two concepts of a theory T is a special case adding an axiom to T. The converse is also true; adding an axiom φ to T is equivalent to collapsing the concept φ with any theorem of T. So using the symmetric closure of the above relation for generating a distance will give the axiomatic distance (Definition 3.6).

6.1. Dropping symmetry.

In several cases, it might be natural not to assume the symmetry of distances between theories. For example, any inconsistent theory is understood to be of axiomatic distance 1 from any consistent theory; we just need to add a contradiction as an axiom. But starting from an inconsistent theory, we can never reach a consistent theory by adding axioms; so considering this distance to be ∞ seems more natural.

Now, let us mimic the work of section 3 under the consideration that symmetry is not required. For instance, a ***directed cluster network*** is a triple $(\mathrm{X}, \mathrm{E}, \mathrm{R})$, where (X, E) is a cluster and R is an arbitrary relation on X. For directed cluster network $\mathcal{X} = (\mathrm{X}, \mathrm{E}, \mathrm{R})$, the ***directed step distance***

$\vec{\mathrm{d}}_{\mathcal{X}} : \mathrm{X} \times \mathrm{X} \to \mathbb{N} \cup \{\infty\}$ can be defined completely analogously to step distance. With a very similar argument to the proof of Theorem 3.4, we can see that the following are true: For each $T_1, T_2, T_3 \in \mathrm{X}$,

(a) $\vec{\mathrm{d}}_{\mathcal{X}}(T_1, T_2) \geq 0$ and $\vec{\mathrm{d}}_{\mathcal{X}}(T_1, T_2) = 0 \iff T_1 \mathrel{\mathrm{E}} T_2$.
(b) $\vec{\mathrm{d}}_{\mathcal{X}}(T_1, T_2) \leq \vec{\mathrm{d}}_{\mathcal{X}}(T_1, T_3) + \vec{\mathrm{d}}_{\mathcal{X}}(T_3, T_2)$.

Definition 6.3. Recall the axiom adding relation $\leftarrow$ introduced on page 7 herein. Let X be a class of theories, then $(\mathrm{X}, \equiv, \leftarrow)$ is a directed cluster network. It is directed step distance is called the ***directed axiomatic distance*** *on* X.

Apparently, Theorem 3.7 is no longer true if we replace the axiomatic distance by the mimicked directed one. So, this is one of the situations where dropping the symmetry might be more interesting. Not just the axiomatic distance can be directed, but also the conceptual distance, interpretation distance, and so on.

It might be more appropriate to call these distance "uni-directed distances", indeed in the directed cluster network the single steps are determined by only one relation. We can also define *bi-directed distances* or *multi-directed distances* where the single steps can be determined by two or more relations.

For example, it might be useful to introduce the a "new conceptual distance" that measures the minimum number of concepts needed to be added to or removed from one theory to reach the other theory up to definitional equivalence. We already have a notion for concept adding (recall $\rightsquigarrow$ in Definition 4.1). A precise definition for concept removal is also required. So, in this case, two relations will indicate the single steps and such new conceptual distance must be a bi-directed step distance.

The notion of concept-removal below is inspired by the idea how the concept of faster-than-light observers were removed from the theory capturing classical kinematics in [Lefever, 2017] and [Lefever and Székely, 2018a].

Definition 6.4. Let T, T^- be two theories. We say that T^- is a ***concept-removal*** *of* T and we write $T \gtrdot T^-$ iff there is $\varphi \in \mathsf{Fm}$ such that $T^- = T_m \cup \{\neg\varphi\}$ for some maximal consistent subtheorey of $\mathsf{Cn}(T)$ for which $T_m \not\models \varphi$.

Let us also note that, in some cases, it is possible to remove finitely many concepts in only one step. For instance, recall from section 4, the propositional theories P_n's. We have shown there that we need exactly two steps of adding concepts to get P_4 from P_2. However, removing a single concept, namely $(c_2(x) \neq c_3(x)) \vee (c_3(x) \neq c_4(x))$, from P_4 gives P_2.

Definition 6.5. Let X be a class of theories and let R be the union of the relations $\rightsquigarrow$ and $\gtrdot$. The ***bi-directed conceptual distance*** *on* X is the directed step distance of the directed cluster network $(\mathrm{X}, \rightleftarrows, \mathrm{R})$. In the case when X is the class of all theories, we denote this distance by $\vec{\mathrm{C}}\mathrm{d}$.

The facts $\vec{\mathrm{C}}\mathrm{d}(P_2, P_4) = 2$ and $\vec{\mathrm{C}}\mathrm{d}(P_4, P_2) = 1$ show that the bi-directed distance is not symmetric. Hence conceptual and bi-directed conceptual distances are different, but it is still interesting to understand how much different they are. For example, the following problem is worth investigating.

Problem 6. Are there two theories T and T' such that

$$\mathrm{Cd}(T, T') \neq \vec{\mathrm{C}}\mathrm{d}(T, T') \quad \text{and} \quad \mathrm{Cd}(T, T') \neq \vec{\mathrm{C}}\mathrm{d}(T', T)?$$

It is not straightforward to find this example, each conceptual distance calculated in the present paper coincides with one of its corresponding bi-directed conceptual distances. In order to do the same with axiomatic distance, i.e., to define a *bi-directed axiomatic distance*, all we need is a precise definition for theorem removal. We propose the following theorem-removal notion.

Definition 6.6. Let T, T^- be two theories. We say that T^- is a ***theorem-removal*** *of* T iff there is $\varphi \in \mathsf{Fm}$ such that $T \models \varphi$ and $T^- = T_m$, where T_m is maximal consistent subtheorey of $\mathsf{Cn}(T)$ for which $T_m \not\models \varphi$.

7. Concluding philosophical remarks

One very important topic in the philosophy of science, is how different scientific theories can be compared to each other, especially in the case of competing theories. The first criterion for theory comparison is empirical adequacy. One theory is better than another if it accounts for more of the data or phenomena than another. In the past, this has often been, or has been presented as being, a fairly straightforward matter to decide by philosophers and historians.

This is oversimplified for two reasons. One is that sometimes in the history of science, one theory accounts for some of the data or phenomena very well. Another accounts for another area of data very well. They both agree on, but have different accounts of, the same data and both have failings. The comparison of two theories in terms of empirical adequacy requires that we count the data or the phenomena. Deciding what to count, and how to assign weight to it, has some arbitrariness to it. This is very well illustrated in [Chang, 2012] where he discusses the history of the competition between the phlogiston theory of water and the compound theory of water. With Chang, we conclude that deciding that one theory is more empirically adequate than another is not at all times, and in all circumstances, simple and straightforward, and with the fragmentation of science into more and more specialized areas of research, it is increasingly rare to find empirical adequacy to be enough to decide between competing theories.

Worse: with more complicated and "cutting-edge" examples, we find that observation in science is not simple, but is informed by theory and language; making an observation is an informed and educated act. At the edges of science, we make observations using highly specialized instruments, which are constructed based on their own theories. Thus, what looked in the past to be a relatively simple judgment to make: "this scientific theory is better than this other" turns out to be rather subtle; since it requires individuation and assigning weights, in a way that is arbitrary with respect to the theories themselves.

Under an over simplified view of the unity of science, the subtlety threatens any pretense science has to objectivity, because what counts as a true and verifiable statement takes specialized instruments that we assume to work according to the theory we have of the instrument. If someone has an alternative account, then the explanation for the phenomenon changes. As a result, if we want to recover some semblance of objectivity in science, it is ever more pressing to receive confirmation of a theory from other directions outside the theory.

In terms of objectivity, one reassuring feature of science is its precision. Logic is the most precise form of investigation. Under the pressure of our considerations above, when we have several logical theories that are each to some extent empirically adequate, it is not clear that we should retain one and dismiss the other. We then have pluralism in science. Pluralism in science is an obvious philosophical position when we consider that several theories are all more-or-less empirically adequate, and show merits with respect to other, incomparable, or only artificially comparable, virtues and vices. The virtues might include: simplicity (determined by language, proofs, metaphysical parsimony or concepts), meeting a particular goal of the scientists, neatness of categorization, breadth of categorisation, standardization of explanation or meeting operational opportunities and so on. Vices might include: complexity (determined by language, proofs, metaphysical elaborations that have little use within a theory or concepts), goal failure, messy or narrow categorization, non-standard explanations, being too ambitious and not going outside the constraints of the operations available. With this plethora of incommensurable theories all competing and each adequate in their own way and for their own purpose, we look elsewhere than between the theories in and of themselves and the data to make sense of the present state of science. What we then look for are other ways of comparing theories, while accepting them until such time as we come up against a good reason to give one up, such as: its being refuted by new evidence or being too remote from too many other theories to be worth pursuing (now).

For this reason, the relations between theories, independent of their relation to reality, becomes very important. Until now, this area of study has mostly been *qualitative*. In the present paper, we explore a new *quantitative* approach: the measured distance between theories. To establish this distance, we need to study the structure of the differences, i.e., the connections, between theories. By developing several such metrics based on our definitions, and noticing that some are less interesting than others, we already learn a lot. Counting axioms does not give us much information about the distance between theories. Counting concepts is much more subtle and informative.

One area of study that has a close relationship with the notion of conceptual distance is that of complexity. As we know, complexity, also can be measured in several ways: Turing complexity, in terms of the analytic hierarchy, and so on. If one theory is more complex than another in one of these measures, then it is natural to investigate the relationship between that and the distances we look at here. Some of the significance of the present work might be in its relationship to complexity theory. This is a subject of future investigation.

The idea of having a notion of distance between theories (of the same nature) seems applicable in any science. In computer science, programming languages and other systems can be seen as axiomatized theories. For more details about this, see, e.g., [Floyd, 1967], [Hoare, 1969] and [Meyer and Halpern, 1982]. Hence, it seems also natural to search for the best fit notion of equivalence between these theories. Developing this may give us insight to determine what can be one step difference between such two theories. Having these in hand, a distance can be then defined in the same way of section 3 herein. The novelty here would be in choosing such equivalence and one step relation in a way guarantees that the corresponding step distance is applicable.

Acknowledgement.

We shall thank Hajnal Andréka, Jean Paul Van Bendegem, Zalán Gyenis, Péter Juhász, Judit X. Madarász, István Németi and others for contributing in this work by their fruitful comments and questions.

References

[Andréka et al., 2004] Andréka, H., Madarász, J. X., and Németi, I. (2004). Logical analysis of relativity theories. In *First-order Logic Revisited*, pages 7–36. Logos Verlag.

[Andréka et al., 2005] Andréka, H., Madarász, J. X., and Németi, I. (2005). Mutual definability does not imply definitional equivalence, a simple example. *Mathematical Logic Quarterly*, 51,6:591–597.

[Andréka et al., 2012] Andréka, H., Madarász, J. X., Németi, I., and Székely, G. (2012). A logic road from special relativity to general relativity. *Synthese*, 186,3:633–649.

[Andréka et al., 2002] Andréka, H., Madarász, J. X., Németi, I., with contributions from: Andai, A., Sági, G., Sain, I., and Tőke, C. (2002). *On the logical structure of relativity theories*. Research report, Alfréd Rényi Institute of Mathematics, Hungar. Acad. Sci., Budapest. https://old.renyi.hu//pub/ algebraic-logic/Contents.html.

[Baltag and Smets, 2005] Baltag, A. and Smets, S. (2005). Complete axiomatizations for quantum actions. *International Journal of Theoretical Physics*, 44(12):2267–2282.

[Barrett and Halvorson, 2016] Barrett, T. W. and Halvorson, H. (2016). Glymour and quine on theoretical equivalence. *Journal of Philosophical Logic*, 45(5):467–483.

[Carathéodory, 1909] Carathéodory, C. (1909). Untersuchungen über die grundlagen der thermodynamik. *Math. Ann.*, 64:355–386.

[Chang, 2012] Chang, H. (2012). *Is Water H_2O? Evidence, Realism and Pluralism*. Springer, Dordrecht.

[Cooper, 1967] Cooper, J. L. B. (1967). The foundations of thermodynamics. *Journal of Mathematical Analysis and Applications*, 17:172–193.

[Floyd, 1967] Floyd, R. W. (1967). Assingining meanings to programs. *Proceedings of symposia in applied mathematics*, 19.

[Henkin et al., 1985] Henkin, L., Monk, J., and Tarski, A. (1985). *Cylindric Algebras Part II*. North-Holland.

[Hoare, 1969] Hoare, C. A. R. (1969). An axiomatic basis for computer programming. *Communications of the ACM*, 12(10):576–583.

[Japaridze and Jongh, 1998] Japaridze, G. and Jongh, D. (1998). The logic of provability. In *Handbook of proof theory, volume 137*, pages 475–550. Elsevier.

[Krause and Arenhart, 2017] Krause, D. and Arenhart, J. R. B. (2017). *The Logical Foundations of Scientific Theories: Languages, Structures, and Models*. Routledge.

[Lefever, 2017] Lefever, K. (2017). *Using Logical Interpretation and Definitional Equivalence to compare Classical Kinematics and Special Relativity Theory*. PhD thesis, Vrije Universiteit Brussel.

[Lefever and Székely, 2018a] Lefever, K. and Székely, G. (2018a). Comparing classical and relativistic kinematics in first-order-logic. *Logique et Analyse*, 61(241):57–117.

[Lefever and Székely, 2018b] Lefever, K. and Székely, G. (2018b). On generalization of definitional equivalence to languages with non-disjoint signatures. pre-print, arXiv:1802.06844.

[Lieb and Yngvason, 2000] Lieb, E. H. and Yngvason, J. (2000). A fresh look at entropy and the second law of thermodynamics. *Physics Today*, pages 32–37.

[Meyer and Halpern, 1982] Meyer, A. R. and Halpern, J. Y. (1982). Axiomatic def'mitions of programming languages: A theoretical assessment. *Journal of the Association for Computing Machinery*, 29(2):555–576.

[Pinter, 1978] Pinter, C. C. (1978). Properties preserved under definitional equivalence and interpretations. *Zeitschr. f. math. Logik und Grundlagen d. nlath.*, 24:481–488.

[Visser, 2006] Visser, A. (2006). Categories of theories and interpretations. In *Logic in Tehran. Proceedings of the workshop and conference on Logic, Algebra and Arithmetic, held October 18–22, 2003, volume 26 of Lecture Notes in Logic*, pages 284–341. ASL, A.K. Peters, Ltd., Wellesley, Mass.

MICHÈLE FRIEND
Columbian College of Arts & Sciences
George Washington University
E-mail: michele@gwu.edu

MOHAMED KHALED
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
E-mail: khaled.mohamed@renyi.mta.hu

KOEN LEFEVER
Centre for Logic and Philosophy of Science
Vrije Universiteit Brussel
E-mail: koen.lefever@vub.be

GERGELY SZÉKELY
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
E-mail: szekely.gergely@renyi.mta.hu