On Inconsistency and Unsatisfiability

Erwin R. Catesbeiana

University of Bremen, Email: erwin@informatik.uni-bremen.de

Abstract. We study inconsistency and unsatisfiability and their relation to soundness, completeness, paraconsistency and conservative extension in an arbitrary logical system (formalised as institution equipped with an entailment system).

1 Introduction

The study of logical inconsistencies has a long tradition that goes back to at least Aristotle. Indeed, Aristotle examined contemporary philosophical arguments and revealed inconsistencies. For example, Anaxagoras imagined the mind to be the initiating principle of all things, and suspending on its axis the balance of the universe; affirming, moreover, that the mind is a simple principle, unmixed, and incapable of admixture, he mainly on this very consideration separates it from all amalgamation with the soul; and yet in another passage he actually incorporates it with the soul. This inconsistency was pointed out by Aristotle, but it remains unclear whether he meant his criticism to be constructive, and to fill up a system of his own, rather than destructive of the principles of others.[1]

Aristotle himself created a rich source of what perhaps not can be called inconsistencies, but false theorems: a number of his Syllogisms exhibit the *existential fallacy*, i.e. have implicit existential assumptions, which means that they are unsound when read literally. For example, the *Fesapo* syllogism:

$$\frac{\begin{array}{c}\text{No humans are perfect.}\\ \text{All perfect creatures are mythical.}\end{array}}{\text{Some mythical creatures are not human.}}$$

After Aristotle came a long period of logical desert, with only Scholastic arguments, which only was left in the 19th century with the discoveries of Boole, Frege and others. Actually, Frege created a rich and powerful logical system called "Begriffsschrift". It was discovered to be inconsistent by Russell in the early 20th century, by a proof that resembles the barber paradox: assume that there is a town with a barber that shaves all people who do not shave themselves. Then the barber shaves himself iff he does not — a contradiction.

The origin of this inconsistency is the power of self-application, i.e. the possibility to apply predicates to themselves. For example, *monosyllabic* is an adjective that does not hold of itself, where as *polysyllabic* does. Now let *non-self-referential* be the adjective expressing that an adjective does not hold for itself.

[1] See `http://www.ccel.org/ccel/schaff/anf03.iv.xi.xii.html`.

That is, monosyllabic is non-self-referential (and polysyllabic isn't). Is non-self-referential non-self-referential? It is the merit of the modern web ontology language OWL-full to have provided, in the early 21st century and more than 120 years after Frege, a logic where again predicates can be applied to themselves.

In the modern use of logic in theoretical computer science, the notion of inconsistency (and its companion unsatisfiability) gains importance in the area for formal specification and software development with formal methods — our jubilarian Bernd Krieg-Brückner has been quite active in this research area. A central idea is that in an early phase of the development process, initial requirements are first formulated informally and then are formalised, such that intended logical consequences can be checked, and inconsistencies (that prevent the development of a correct implementation) can be found. And indeed, not only programs are notoriously buggy, but also their specifications tend to be incorrect and inconsistent. Modern specification languages like CafeOBJ and CASL come with libraries of specifications that also provide examples of such inconsistencies. Indeed, these languages are feature-rich and complex, which eases the development of non-trivial[2] inconsistent theories. In this context, we should also mention the research on upper ontologies, which are usually quite large and inconsistent first-order theories.[3] Indeed, in the SUMO ontology, several inconsistencies have been found [5]. The SUMO $100 Challenges[4] explicitly call for demonstrating the consistency or inconsistency of (parts of) SUMO. Needless to say that so far, only inconsistencies have been found.

2 Institutions and Logics

The study of inconsistency and unsatisfiability can be carried out largely independently of the nature of the underlying logical system. We use the notion of *institution* introduced by Goguen and Burstall [3] in the late 1970ies. It approaches the notion of logical system from a relativistic view: rather than treating the concept of logic as eternal and given, it accepts the need for a large variety of different logical systems, and ask for the central principles of logical systems. They arrived at a mathematical formalisation of that notion which is really ingenious and lead to rich and flourishing theory that is laid out in hundreds of research papers, and most recently also in a book [1].

While the notion of institution takes a model.theoretic perspective, it was later complemented by the more proof-theoretic notion of entailment system (also called Π-institution) [8, 2].

What are the essential ingredients of a logical system? It surely has a notion of *sentence*, and derivability relation $\vdash$ on sentences that allows to derive conclusions from a given set of assumptions. Moreover, at the model-theoretic

[2] Of course, the set of logical consequences of an inconsistent theory is always trivial. However, the axioms of the theory itself may have varying complexity and subtlety of interaction in order to lead to the inconsistency.

[3] Although there are more efficient ways of writing up such theories ...

[4] See `http://www.cs.miami.edu/~tptp/Challenges/SUMOChallenge/`.

side, there is a notion of *model* and a *satisfaction relation* between models and sentences, denoted by $\models$. The latter leads to the relation of *logical consequence* (all models satisfying the premises also satisfy the conclusion); this relation is also denoted by $\models$. A logic is *sound*, if $\Gamma \vdash \varphi$ implies $\Gamma \models \varphi$, and *complete*, if the converse implication holds.

Moreover, an important observation is that all this structure depends on the context, i.e. the set of non-logical (or user-defined) symbols around. These contexts are called *signatures* and formalised just as object of an abstract category. The reader not familiar with category theory or not interested in the formal details can safely skip the rest of this section and just keep in mind the above informal motivations.

Definition 1. *An* entailment system *consists of*

- *a category* $\mathbb{S}ign^{\mathcal{E}}$ *of signatures and signature morphisms,*
- *a functor* $\mathsf{Sen}^{\mathcal{E}}\colon \mathbb{S}ign^{\mathcal{E}} \longrightarrow \mathbb{S}et$ *giving, for each signature* Σ*, the set of* sentences $\mathsf{Sen}^{\mathcal{E}}(\Sigma)$*, and for each signature morphism* $\sigma\colon \Sigma \longrightarrow \Sigma'$*, the* sentence translation map $\mathsf{Sen}^{\mathcal{E}}(\sigma)\colon \mathsf{Sen}^{\mathcal{E}}(\Sigma) \longrightarrow \mathsf{Sen}^{\mathcal{E}}(\Sigma')$*, where often* $\mathsf{Sen}^{\mathcal{E}}(\sigma)(\varphi)$ *is written as* $\sigma(\varphi)$*,*
  $\vdash_\Sigma \subseteq \mathcal{P}(\mathsf{Sen}(\Sigma)) \times \mathsf{Sen}(\Sigma)$ *for each* $\Sigma \in |\mathbb{S}ign|$*, such that the following properties are satisfied:*
  1. reflexivity: *for any* $\varphi \in \mathsf{Sen}(\Sigma)$*,* $\{\varphi\} \vdash_\Sigma \varphi$*,*
  2. monotonicity: *if* $\Psi \vdash_\Sigma \varphi$ *and* $\Psi' \supseteq \Psi$ *then* $\Psi' \vdash_\Sigma \varphi$*,*
  3. transitivity: *if* $\Psi \vdash_\Sigma \varphi_i$ *for* $i \in I$ *and* $\Psi \cup \{\varphi_i \mid i \in I\} \vdash_\Sigma \psi$*, then* $\Psi \vdash_\Sigma \psi$*,*
  4. $\vdash$-translation: *if* $\Psi \vdash_\Sigma \varphi$*, then for any* $\sigma\colon \Sigma \longrightarrow \Sigma'$ *in* $\mathbb{S}ign$*,* $\sigma(\Psi) \vdash_{\Sigma'} \sigma(\varphi)$*.*
- *for each signature* $\Sigma \in |\mathbb{S}ign^{\mathcal{E}}|$*, an entailment relation* $(\mathsf{Sen}(\Sigma), \vdash^{\mathcal{E}}_{\Sigma})$*,*
- *for each signature morphism* $\sigma\colon \Sigma_1 \longrightarrow \Sigma_2$*, a sentence translation map* $\mathsf{Sen}^{\mathcal{E}}(\Sigma_1) \to \mathsf{Sen}^{\mathcal{E}}(\Sigma_2)$ *preserving* $\vdash^{\mathcal{E}}$*. By abuse of notation, we will also denote this map with* σ*.*

A *theory* is a pair (Σ, Γ) where Γ is a set of Σ-sentences. An *entailment theory morphism*[5] $(\Sigma, \Gamma) \longrightarrow (\Sigma', \Gamma')$ is a signature morphism $\sigma\colon\ \Sigma \longrightarrow \Sigma$ such that $\Gamma' \vdash_{\Sigma'} \sigma(\Gamma)$.

Let $\mathbb{C}at$ be the category of categories and functors.[6]

Definition 2. *An* institution $\mathcal{I} = (\mathbb{S}ign^{\mathcal{I}}, \mathsf{Sen}^{\mathcal{I}}, \mathsf{Mod}^{\mathcal{I}}, \models^{\mathcal{I}})$ *consists of*

- *a category* $\mathbb{S}ign^{\mathcal{I}}$ *of* signatures*,*
- *a functor* $\mathsf{Sen}^{\mathcal{I}}\colon \mathbb{S}ign^{\mathcal{I}} \longrightarrow \mathbb{S}et$ *(as for entailment systems),*

[5] also called *interpretation of theories*.

[6] Strictly speaking, $\mathbb{C}at$ is not a category but only a so-called quasicategory, which is a category that lives in a higher set-theoretic universe [4]. However, we do not care about this here. Indeed, these foundational questions are ignored by most mathematicians and computer scientists, since ignoring them provides a rich source of inconsistencies.

– *a functor* $\mathsf{Mod}^{\mathcal{I}}\colon (\mathbb{S}ign^{\mathcal{I}})^{op} \longrightarrow \mathbb{C}at$ *giving, for each signature* Σ*, the category of* models $\mathsf{Mod}^{\mathcal{I}}(\Sigma)$*, and for each signature morphism* $\sigma\colon \Sigma \longrightarrow \Sigma'$*, the* reduct functor $\mathsf{Mod}^{\mathcal{I}}(\sigma)\colon \mathsf{Mod}^{\mathcal{I}}(\Sigma') \longrightarrow \mathsf{Mod}^{\mathcal{I}}(\Sigma)$*, where often* $\mathsf{Mod}^{\mathcal{I}}(\sigma)(M')$ *is written as* $M'|_\sigma$*,*
– *a satisfaction relation* $\models^{\mathcal{I}}_{\Sigma} \subseteq |\mathsf{Mod}^{\mathcal{I}}(\Sigma)| \times \mathsf{Sen}^{\mathcal{I}}(\Sigma)$ *for each* $\Sigma \in \mathbb{S}ign^{\mathcal{I}}$*,*

such that for each $\sigma\colon \Sigma \longrightarrow \Sigma'$ *in* $\mathbb{S}ign^{\mathcal{I}}$ *the following* satisfaction condition *holds:*

$$M' \models^{\mathcal{I}}_{\Sigma'} \sigma(\varphi) \leftrightarrow M'|_\sigma \models^{\mathcal{I}}_{\Sigma} \varphi$$

for each $M' \in \mathsf{Mod}^{\mathcal{I}}(\Sigma')$ *and* $\varphi \in \mathsf{Sen}^{\mathcal{I}}(\Sigma)$*.* □

We will omit the index $\mathcal{I}$ when it is clear from the context.

A *logic* is an entailment system paired with an institution agreeing on their signature and sentence parts. Usually, a logic is required to be *sound*, that is, $\Gamma \vdash_\Sigma \varphi$ implies $\Gamma \models_\Sigma \varphi$. If also the converse holds, the logic is *complete*.

A *theory* is defined as for entailment systems. An *institution theory morphism* $(\Sigma, \Gamma) \longrightarrow (\Sigma', \Gamma')$ is a signature morphism $\sigma :\ \Sigma \longrightarrow \Sigma$ such that $\Gamma' \models_{\Sigma'} \sigma(\Gamma)$. Let $\mathbb{T}h(\mathcal{I})$ denote this category. Each theory (Σ, Γ) inherits sentences from $\mathsf{Sen}^{\mathcal{I}}(\Sigma)$, while the models are restricted to those models in $\mathsf{Mod}^{\mathcal{I}}(\Sigma)$ that satisfy all sentences in Γ. It is easy to see that $\mathcal{I}$ maps theory morphisms to corridors in this way. By taking $\mathbb{T}h(\mathcal{I})$ as "signature" category, we arrive at the institution $\mathcal{I}^{Th}$ of theories.

Example 3. Classical propositional logic (**CPL**) has the category $\mathbb{S}et$ of sets and functions as its signature category. Σ-sentences are given by the following grammar

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \to \varphi_2 \mid \top \mid \bot$$

where p denotes propositional variables from Σ. Sentence translation is the obvious replacement of propositional variables.

Σ-models are functions from Σ to $\{T, F\}$. A (unique) model morphism between two models exists if the values of variables change only from F to T, but not vice versa. Given a signature morphism $\sigma\colon \Sigma_1 \longrightarrow \Sigma_2$, the σ-reduct of a Σ_2-model $M_2\colon \Sigma' \longrightarrow \{T, F\}$ is given by composition: $M_2 \circ \sigma$. This obviously extends to reducts of morphisms.

Satisfaction is inductively defined by the usual truth table semantics. Since reduct is composition, it is straightforward to prove the satisfaction condition.[7]

The entailment relation is the minimal relation satisfying the properties listed in Table 1, plus the property $\neg\neg\varphi \vdash \varphi$.

This logic is sound and complete. □

[7] A precise argument is as follows: the Boolean operations form a signature such that Sen is the free algebra functor for algebras over that signature (these algebras are Boolean algebras without laws). $\{T, F\}$ also is such an algebra, denoted by $Bool$, and for a model M, sentence evaluation is $\varepsilon_{Bool} \circ \mathsf{Sen}(M)$, where ε is the counit. Then the satisfaction condition is $\varepsilon_{Bool} \circ \mathsf{Sen}(M) \circ \mathsf{Sen}(\sigma) = \varepsilon_{Bool} \circ \mathsf{Sen}(M \circ \sigma)$, which is just functoriality of Sen.

A Heyting algebra H is a partial order $(H, \leq)$ with a greatest element $\top$ and a least one $\bot$ and such that any two elements $a, b \in H$

- have a greatest lower bound $a \wedge b$ and a least upper bound $a \vee b$, and
- there exists a greatest element x such that $a \wedge x \leq b$; this element is denoted $a \Rightarrow b$.

In a Heyting algebra, we can define a derived operation $\neg$ by $\neg a := a \Rightarrow \bot$. A Heyting algebra morphism $h\colon H_1 \longrightarrow H_2$ is a map preserving all the operations.

Example 4. Heyting-algebra based intuitionistic propositional logic (**IPL-HA**) inherits the signature category and sentences from **CPL**.

A Σ-model (ν, H) consist of a Heyting algebra H together with a valuation function $\nu\colon \Sigma \longrightarrow |H|$ into the underlying set $|H|$ of H. Σ-model morphisms $h\colon (\nu_1, H_1) \longrightarrow (\nu_2, H_2)$ are Heyting algebra morphisms $h\colon H_1 \longrightarrow H_2$ such that $h_1 \circ \nu_1(p) \leq \nu_2(p)$ for each $p \in \Sigma$. Again, model reduct is defined by composition, and this easily extends to morphisms.

Using the Heyting algebra operations, it is straightforward to extend the valuation ν of a Σ-model (ν, H) from propositional variables to all sentences: $\nu^{\#}\colon Sen(\Sigma) \longrightarrow |H|$. (ν, H) satisfies a sentence φ iff $\nu^{\#}(\varphi) = \top$. The satisfaction condition follows similarly as for **CPL**.

The entailment relation is the minimal relation satisfying the properties listed in Table 1. This turns **IPL-HA** into a sound and complete logic. □

3 Logical connectives

Within an abstract logic, it is possible to define logical connectives purely be their proof-theoretic and model-theoretic properties. We start with proof theory and adapt the standard definitions from [7].

connective	defining property
proof-theoretic conjunction $\wedge$	$\Gamma \vdash \varphi \wedge \psi$ iff $\Gamma \vdash \varphi$ and $\Gamma \vdash \psi$
proof-theoretic disjunction $\vee$	$\varphi \vee \psi, \Gamma \vdash \chi$ iff $\varphi, \Gamma \vdash \chi$ and $\psi, \Gamma \vdash \chi$
proof-theoretic implication $\rightarrow$	$\Gamma, \varphi \vdash \psi$ iff $\Gamma \vdash \varphi \rightarrow \psi$
proof-theoretic truth $\top$	$\Gamma \vdash \top$
proof-theoretic falsum $\bot$	$\bot \vdash \varphi$
proof-theoretic negation $\neg$	$\Gamma, \varphi \vdash \bot$ iff $\Gamma \vdash \neg\varphi$

Table 1. Properties of proof-theoretic connectives

Note that these properties characterise connective essentially by their *proof-theoretic* behaviour; they mostly even directly correspond to proof rules. Below, we will also introduce *semantic* connectives. A logic is said to *have a proof-theoretic connective* if it is possible to define an operation on sentences with the properties specified in Table 1. For example, both **IPL-HA** and **CPL** have all proof-theoretic connectives.

We also can define internal connectives at the semantic level. A logic is said to *have a semantic connective* if it is possible to define an operation on sentences with the specified properties.

connective	defining property
semantic disjunction $\vee$	$M \models \varphi \vee \psi$ iff $M \models \varphi$ or $M \models \psi$
semantic conjunction $\wedge$	$M \models \varphi \wedge \psi$ iff $M \models \varphi$ and $M \models \psi$
semantic implication $\rightarrow$	$M \models \varphi \rightarrow \psi$ iff $M \models \varphi$ implies $M \models \psi$
semantic negation $\neg$	$M \models \neg\varphi$ iff $M \not\models \varphi$
semantic truth $\top$	$M \models \top$
semantic falsum $\bot$	$M \not\models \bot$

Table 2. Properties of semantic connectives

While **CPL** has all semantic connectives (indeed, they coincide with the proof-theoretic ones), **IPL-HA** only has semantic conjunction, truth and falsum.

4 Inconsistency and Unsatisfiability

In the sequel, let us fix a logic $\mathcal{L}$ (in the above sense), which a priori need neither be sound nor complete.

The notion of unsatisfiability is quite clear:

Definition 5. *A theory is unsatisfiable if it has no models.*

Indeed, the notion of inconsistency is more interesting than unsatisfiability[8], since it enjoyed some historical development. According to Aristotle, inconsistency means that both some sentence as well as its negation can be proved.

Definition 6. *Assume that $\mathcal{L}$ has a negation connective (not necessarily being a proof-theoretic or semantic negation[9]), and let T be a theory in $\mathcal{L}$. T is* Aristotle inconsistent*, if there is some sentence φ with $T \vdash \varphi$ and $T \vdash \neg\varphi$.*

This notion has several disadvantages. Firstly, it presupposes the notion of negation, which is not available in all logics. More importantly, it classifies paraconsistent logics as inconsistent, which can only be apologised by the fact that paraconsistent logics were not known at Aristotle's time.

[8] Although the Rolling Stones devoted a song to unsatisfiability.

[9] However, let us assume that if there is a proof-theoretic negation, then this is used. Otherwise, the notion of inconsistency of course depends on the chosen connective.

A modern definition of inconsistency overcoming this problem was coined by David Hilbert. Hilbert was the founder of the famous "Hilbert programme", the aim of which was to prove the consistency of all of mathematics by reducing it to the consistency of a small number of finitary principles, for which there is enough faith into their consistency. Hilbert's programme greatly failed, as was shown by Gödel's second incompleteness theorem (actually, the name is misleading: it should be called Gödel's Great Inconsistency Theorem):

Theorem 7 (Gödel). *There is a first-order theory T of zero, successor, addition and ordering on the natural numbers (which is actually quite simple and weak), such that for any extension T' of T, if T' can prove its own consistency (encoded as a statement on natural numbers), then T' is inconsistent.*[10]

Hence, although Hilbert's programme was a powerful and striking idea, in the end it could not be successful. As a result, the question whether the theories like ZFC (that are used as a foundation of mathematics and theoretical computer science!) are consistent or inconsistent is open. Indeed, the only way to firmly resolve this open question would be to prove the inconsistency of ZFC.[11] All what we have so far are relative results, such as the famous result by Gödel:

ZFC is inconsistent iff ZF is inconsistent.[12]

which means that when looking for an inconsistency proof for ZF, we equally well may use the stronger (and hence easier to prove inconsistent) system ZFC.

But even though Hilbert's programme failed, Hilbert left us with a modern definition of inconsistency:

Definition 8 (Hilbert). *Assume that $\mathcal{L}$ has a falsum constant (not necessarily being a proof-theoretic or semantic falsum*[13]*). T is* $\bot$-inconsistent*, if $T \vdash \bot$.*

Still, this definition does not work with logics that do not have a falsum, for example positive or equational logic. Hilbert therefore also found a notion of inconsistency that has the least prerequisites (no logical connectives are needed) and that simultaneously is most powerful (in terms of logical strength of inconsistent theories):

Definition 9 (Hilbert). *T is* absolutely inconsistent*, if $T \vdash \varphi$ for any sentence φ of same signature as T. This is also known as the principle* ex contradictione quodlibet *(from a contradiction, one can follow everything).*

---

[10] For first-order logic, the various notions of inconsistency discussed so far are equivalent; hence we can be unspecific here.

[11] For an initial attempt in this direction, see [6].

[12] Actually, Gödel proved the corresponding statement about unsatisfiability, but by Gödel's completeness theorem for first-order logic, inconsistency and unsatisfiability are equivalent, see our Prop. 12 below.

[13] Again, let us assume that if there is a proof-theoretic falsum, then this is used. Otherwise, the notion of inconsistency depends on the chosen constant.

We should also mention a notion of inconsistency introduced by Emil Post: a propositional theory T is Post-inconsistent, if it can derive a propositional variable (not occurring in the axioms of T; the signature possibly needs to be enlarged to get such a variable). Unfortunately, this notion is too much tied to a specific logical system to be of interest here.

The different notions of inconsistency[14] are related as follows:

Proposition 10. *1. absolute inconsistency implies $\perp$-inconsistency and Aristotle inconsistency.*

2. In presence of proof-theoretic falsum, absolutely inconsistency and $\perp$-inconsistency are equivalent.

3. In presence of proof-theoretic falsum and negation, all three notions of inconsistency re equivalent.

Proof. 1. obvious

2. Directly from the definition of proof-theoretic falsum.

3. By 1. and 2., it remains to show that Aristotle inconsistency.implies absolutely inconsistency. By definition of proof-theoretic negation, from $\Gamma \vdash \neg\varphi$ we get and $\Gamma \cup \{\varphi\} \vdash \perp$. Together with $\Gamma \vdash \varphi$, this leads to and $\Gamma \vdash \perp$. ☐

5 Soundness and Completeness, with an Application to Paraconsistency

Inconsistency and unsatisfiability also play a great role in determining whether a logic is sound or complete.

We begin with a simple lemma showing that falsum and truth are two sides of the same coin:

Lemma 11. *In presence of proof-theoretic negation, falsum and truth,*

$$\neg\perp \dashv\vdash \top \text{ and } \neg\top \dashv\vdash \perp.$$

Soundness and completeness, while defined in terms of entailment, can be characterised completely in terms of inconsistency and unsatisfiability.

Proposition 12. *Let $\mathcal{L}$ be a logic with both proof-theoretic and semantic negation, truth and falsum. (Then, by Prop. 10, all notions of inconsistency are equivalent, hence we can be unspecific in the sequel.)*

1. A logic is sound iff every inconsistent theory is unsatisfiable.

2. A logic is complete iff every unsatisfiable theory is inconsistent.

[14] We credit `http://home.utah.edu/~nahaj/logic/structures/systems/inconsistent.html` for an excellent overview of these notions.

Proof. (1), "⇒" Let T be inconsistent, i.e. $T \vdash \bot$. By soundness, $T \models \bot$, hence T is unsatisfiable.

(1), "⇐" Let $T \vdash \varphi$, then $T \cup \{\neg\varphi\}$ is inconsistent, hence, by the assumption, also unsatisfiable. But this means that $T \models \varphi$.

(2), "⇒" Let $T \models \varphi$. Then $T \cup \{\neg\varphi\}$ is not satisfiable, hence by the assumption, it is inconsistent. From $T \cup \{\neg\varphi\} \vdash \bot$, we obtain $T \cup \{\neg\varphi\} \vdash \neg\neg\bot$ and hence $T \cup \{\neg\bot\} \vdash \varphi$. By Lemma 11, $T \cup \{\top\} \vdash \varphi$. By the properties of $\top$ and transitivity, we get $T \vdash \varphi$.

(2), "⇐" Let T be unsatisfiable, i.e. $T \models \bot$. By completeness, $T \vdash \bot$, hence T is inconsistent. □

It should be stressed that these proofs become less elegant when reformulated in terms of consistency and satisfiability, as some over-cautious logicians do — they tend to be easily frightened by inconsistencies[15]. The more natural relation is indeed that between inconsistency and unsatisfiability.

Definition 13. *A logic is* paraconsistent *if it admits a theory that is Aristotle inconsistent but absolutely consistent.*

Proposition 14. *1. A paraconsistent logic cannot have proof-theoretic negation.*
2. A sound and complete paraconsistent logic cannot have model-theoretic negation.

Proof. 1. By Prop. 10.

2. Let T be an Aristotle inconsistent theory, i.e. $T \vdash \psi$ and $T \vdash \neg\psi$ for some sentence ψ. By soundness, $T \models \psi$ and $T \models \neg\psi$. By the definition of model-theoretic negation, T is unsatisfiable. But then for any sentence φ, $T \models \varphi$, thus by completeness, also $T \vdash \varphi$. Hence, T is absolutely inconsistent.

6 Conservative Extensions

Definition 15. *A theory morphism* $\sigma\colon T_1 \longrightarrow T_2$ *is* model-theoretically conservative*, if each* T_1*-model has a* σ*-expansion to a* T_2*-model. It is* consequence-theoretically conservative*, if for each sentence* φ *of same signature as* T_1*,*

$$T_2 \models \sigma(\varphi) \text{ implies } T_1 \models \varphi.$$

It is proof-theoretically conservative*, if the same holds for* $\vdash$*, i.e. for each sentence* φ *of same signature as* T_1*,*

$$T_2 \vdash \sigma(\varphi) \text{ implies } T_1 \vdash \varphi.$$

The relation between these notions is as follows:

[15] This goes as far as the Wikipedia website for "Inconsistency" being redirected to "Consistency"!

Proposition 16. *Model-theoretic conservativity implies consequence-theoretic conservativity (but not vice versa). In a sound and complete logic, consequence-theoretic and proof-theoretic conservativity are equivalent.*

Proof. Concerning the first statement, let $T_2 \models \sigma(\varphi)$. We need to show $T_1 \models \varphi$. Let M_1 be a T_1-model. By model-theoretic conservativity, it has an expansion M_2 to a T_2-model; hence also $M_2 \models \sigma(\varphi)$. By the satisfaction condition, $M_1 = M_2|_\sigma \models \varphi$.

The second statement is obvious. □

The importance of conservativity is that it allows to reduce inconsistency (resp. unsatisfiability) of a smaller theory to that of larger one, using the following obvious result:

Proposition 17.
1. *Model-theoretically conservative theory morphisms reflect unsatisfiability.*
2. *Proof-theoretically conservative theory morphisms reflect inconsistency.*

Typically, the target of a conservative theory morphism is larger than (or even an extension of) the source. At first sight, it may sound strange to reduce the goal of showing inconsistency of a given theory to that of showing it for a *larger* one. However, note that generally, larger theories have more axioms, and hence it is more likely to prove that they are inconsistent. (Indeed, the largest theory consisting of *all* sentences is always inconsistent.) Moreover, small inconsistent theories can be boring: who is interested in the smallest inconsistent theory consisting just of falsum? For example, Frege's Begriffsschrift (which is a conservative extension of falsum) is much more interesting.

7 Conclusion

We have summarised and discussed recent research aimed at proving inconsistency of specifications in a structured and institution-independent way. These efforts form part of a larger program aimed at proving the inconsistency of the empty specification, i.e. inconsistency of the meta-framework underlying most of the formalisms under consideration, Zermelo-Fraenkel set theory with choice (ZFC) and hence, by the independence of the axiom of choice as mentioned above, of ZF set theory without choice. In particular, recent research in pure mathematics has been concerned with finding sufficient conditions for the inconsistency of ZF. E.g. it has been shown that the inconsistency of ZF can be reduced to (and is therefore equivalent to) provability in ZF of

- the existence of natural numbers a, b, c and $n \geq 3$ such that $a^n + b^n = c^n$ (Andrew Wiles [13])
- the existence of natural numbers $(a, b, x, y) \neq (2, 3, 3, 1)$ such that $x^a - y^b = 1$ (Preda Mihăilescu [9])
- the existence of a simply connected closed 3-manifold not homeomorphic to S^3 (Grigori Perelman [10–12])

Moreover, ongoing work in automated theorem proving, Thomas Hales' FlysPecK project (http://code.google.com/p/flyspeck/), is directed at reducing the inconsistency of ZF to existence of a sphere packing of average density strictly less than $\pi/18$. In summary, there is good hope that the paradise of mathematics, a widely accepted inconsistent set of foundations, will soon be re-opened.

References

1. R. Diaconescu. *Institution-Independent Model Theory.* Birkhäuser, 2008.
2. J. Fiadeiro and A. Sernadas. Structuring theories on consequence. In D. Sannella and A. Tarlecki, eds., *5th WADT*, vol. 332 of *LNCS*, pp. 44–72. Springer Verlag, 1988.
3. J. A. Goguen and R. M. Burstall. Institutions: Abstract model theory for specification and programming. *Journal of the Association for Computing Machinery*, 39:95–146, 1992. Predecessor in: LNCS 164, 221–256, 1984.
4. H. Herrlich and G. Strecker. *Category Theory.* Allyn and Bacon, Boston, 1973.
5. I. Horrocks and A. Voronkov. Reasoning support for expressive ontology languages using a theorem prover. In *Proceedings of the Fourth International Symposium on Foundations of Information and Knowledge Systems (FoIKS)*, number 3861 in Lecture Notes in Computer Science, pp. 201–218. Springer, 2006.
6. R. E. Kamouna. A paradox ⇒ SAT is (NOT) NP-complete & ZFC is inconsistent. http://arxiv.org/abs/0806.2947.
7. J. Lambek and P. J. Scott. *Introduction to Higher Order Categorical Logic.* Cambridge University Press, 1986.
8. J. Meseguer. General logics. In *Logic Colloquium 87*, pp. 275–329. North Holland, 1989.
9. P. Mihăilescu. Primary cyclotomic units and a proof of catalan's conjecture. *J. Reine angew. Math.*, 572:167–195, 2004.
10. G. Perelman. The entropy formula for the ricci flow and its geometric applications. arXiv:math.DG/, 2002.
11. G. Perelman. Ricci flow with surgery on three-manifolds. arXiv:math.DG/0303109, 2002.
12. G. Perelman. Finite extinction time for the solutions to the ricci flow on certain three-manifolds. arXiv:math.DG/0211159, 2003.
13. A. Wiles. Modular elliptic curves and fermat's last theorem. *Ann. Math.*, 141:443–551, 1995.