Int J Software Informatics, Volume 9, Issue 2 (2015), pp. 141–152
International Journal of Software and Informatics, ISSN 1673-7288

E-mail: ijsi@iscas.ac.cn
http://www.ijsi.org
Tel: +86-10-62661040

On Inconsistency and Unsatisfiability

Till Mossakowski[1] and Lutz Schröder[2]

[1] (Otto-von-Guericke-Universität Magdeburg, Germany)
[2] (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)

Abstract We study inconsistency and unsatisfiability and their relation to soundness, completeness, paraconsistency and conservative extension in generic logical systems (formalized as institutions equipped with an entailment system).
Key words: inconsistency; unsatisfiability; soundness; completeness; paraconsistency; institution

Mossakowski T, Schröder L. On inconsistency and unsatisfiability. *Int J Software Informatics*, **Vol.9, No.2 (2015): 141–152. http://www.ijsi.org/1673-7288/9/i212.htm**

1 Introduction

In the modern use of logic in theoretical computer science, the notion of inconsistency (and its semantic companion, unsatisfiability) gains importance in the area of formal specification and software development with formal methods. A central idea is that in an early phase of the development process, initial requirements are first formulated informally and then are formalized, 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. These languages are feature-rich and complex, which eases the development of non-trivial inconsistent theories[1]. In this context, we should also mention research on upper ontologies, which are usually quite large and inconsistent first-order theories. E.g. a number of inconsistencies have been found in the SUMO ontology by Horrocks and Voronkov[7]. The SUMO $100 Challenges[2] explicitly calls for demonstrating the consistency or inconsistency of (parts of) SUMO. Needless to say that so far, only inconsistencies have been found.

However, the origins of the field are much older. The study of logical inconsistencies has a long tradition that goes back at least to Aristotle. In particular, Aristotle examined contemporary philosophical arguments and revealed inconsistencies contained therein. For example, Anaxagoras imagined "the mind to

[1] 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.
[2] See `http://www.cs.miami.edu/~tptp/Challenges/SUMOChallenge/`.

Corresponding author: Till Mossakowski, Email: mossakow@iws.cs.uni-magdeburg.de
Received 2015-01-12; Revised 2015-03-31; Accepted 2015-04-15.

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) Aristotle has also observed."[3]

Aristotle himself created a rich source of what perhaps cannot 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, like, e.g., 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 there followed a long period of insufficient research into logical inconsistency, interspersed only with some scholastic arguments, which finished as late as in the 19th century with the discoveries of Boole, Frege and others. In particular, Frege created a rich and powerful logical system in his "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 true of itself, whereas *polysyllabic* does. Now let *non-self-referential* be the adjective expressing that an adjective does not hold for itself. 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[4] to have provided, in the early 21st century and more than 120 years after Frege, a logic where predicates can, again, be applied to themselves.

In the present work, we contribute to the field of inconsistency by developing generic, i.e. logic-independent, notions of inconsistency in the framework of institution theory. Not unexpectedly, it turns out that there is a variety of such notions (in particular, Aristotle, Hilbert and absolute inconsistency), which we relate to each other and illuminate in a number of examples (for simplicity, we choose variants of propositional logic here).

2 Institutions and Logics

As indicated in the introduction, 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[5] in the late 1970s. 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 instead asks about common principles shared

[3] Quoted from `http://www.ccel.org/ccel/schaff/anf03.iv.xi.xii.html`.
[4] See `http://www.w3.org/TR/owl-ref`.

across logical systems. For an overview of the development of the theory since its inception, see, e.g., the recent monograph by Diaconescu[3].

While the notion of institution adopts a model-theoretic perspective, it was later complemented by the more proof-theoretic notion of entailment system (also called π-institution)[4,9].

What are the essential ingredients of a logical system? It seems safe to assume that a logical system has a notion of *sentence*, and a syntactic entailment relation $\vdash$ on sentences that allows deriving conclusions from given sets of assumptions. On the model-theoretic side, there is, moreover, a notion of *satisfaction* of sentences by models, typically denoted as an infix satisfaction relation $\models$. 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. These contexts are called *signatures*, and are formalized just as objects 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 2.1. An *entailment system*[9] 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 $\mathsf{Sen}^{\mathcal{E}}(\sigma)(\varphi)$ is often written as $\sigma(\varphi)$,
- for each signature $\Sigma \in |\mathbb{S}ign^{\mathcal{E}}|$, an entailment relation $\vdash_\Sigma \subseteq \mathcal{P}(\mathsf{Sen}(\Sigma)) \times \mathsf{Sen}(\Sigma)$ 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 all $i \in I$ 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 $\mathbb{S}ign$, $\sigma(\Psi) \vdash_{\Sigma'} \sigma(\varphi)$.

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 $\sigma : \Sigma \longrightarrow \Sigma$ such that $\Gamma' \vdash_{\Sigma'} \sigma(\Gamma)$.

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

Definition 2.2. 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[6]. However, we ignore this issue here. Indeed, foundational questions such as this are ignored by most mathematicians and computer scientists, and 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 $\mathsf{Mod}^{\mathcal{I}}(\Sigma)$ of *models*, 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 $\mathsf{Mod}^{\mathcal{I}}(\sigma)(M')$ is often written as $M'|_{\sigma}$, and

– 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 *satisfaction condition*

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

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

We omit the index $\mathcal{I}$ when it is clear from the context. Given a signature Σ, a set $\Gamma \subseteq \mathsf{Sen}(\Sigma)$ of sentences, and a sentence $\varphi \in \mathsf{Sen}(\Sigma)$, we say that φ *is a logical consequence of* Γ, and write $\Gamma \models_{\Sigma} \varphi$, if for all $M \in \mathsf{Mod}(\Sigma)$, $M \models_{\Sigma} \Gamma$ implies $M \models_{\Sigma} \varphi$, where by definition $M \models_{\Sigma} \Gamma$ iff $M \models_{\Sigma} \psi$ for all $\psi \in \Gamma$. Moreover, we write $\Gamma \models_{\Sigma} \Delta$ for $\Delta \subseteq \mathsf{Sen}(\Sigma)$ if $\Gamma \models_{\Sigma} \varphi$ for each $\varphi \in \Delta$.

A *logic* consists of an entailment system and an institution that agree 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 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 the category of theories and institution theory morphisms in $\mathcal{I}$. 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. We can thus define the institution $\mathcal{I}^{Th}$ of theories over $\mathcal{I}$ by taking $\mathbb{T}h(\mathcal{I})$ as the signature category, with the indicated sentence and model functors.

Model morphisms do not matter in the current technical development, so we elide them in the following examples.

Example 2.3. Classical propositional logic (**CPL**) has the category $\mathbb{S}et$ of sets and functions as its signature category, with sets understood as sets of propositional variables. Σ-sentences are given by the grammar

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

where $p \in \Sigma$ is a propositional variable. Sentence translation acts in the obvious way by replacing propositional variables with their images under the signature morphism (which is just a map).

Σ-models are truth valuations, i.e. functions from Σ to $\{T, F\}$. Given a signature morphism $\sigma\colon \Sigma_1 \longrightarrow \Sigma_2$, the σ-reduct of a Σ_2-model $M_2\colon \Sigma_2 \longrightarrow \{T, F\}$ is given by composition: $M_2|_{\sigma} = M_2 \circ \sigma$.

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

[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 like Boolean algebras but without satisfying any equational 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 of the adjunction of algebras over sets. 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.

The entailment relation is the smallest relation satisfying the properties listed in Table 1, plus the property $\neg\neg\varphi \vdash \varphi$ and the usual axiom rule, i.e. $\Gamma \vdash \phi$ whenever $\phi \in \Gamma$. One then, of course, needs to show that this relation is really an entailment relation. Reflexivity is immediate from the axiom rule, and monotonicity and $\vdash$-translation are shown by an easy induction over proofs. Transitivity is non-trivial, as it amounts to cut elimination; for a detailed proof, see, for example, Ref. [14].

This logic is sound and complete.

A Heyting algebra H is a partial order $(H, \leqslant)$ 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 $a \wedge x \leqslant b$; 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 (i.e. $\wedge$, $\vee$, $\Rightarrow$, $\top$, $\bot$).

Example 2.4. 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. Again, model reduct is defined by composition.

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: $\nu^{\#}\colon Sen(\Sigma) \longrightarrow |H|$. Then, (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

In an abstract logic, it is possible to define logical connectives purely by their proof-theoretic and model-theoretic properties. We begin with the proof theoretic approach, and adapt the standard definitions[8]. The defining proof-theoretic properties of some standard connectives are shown in Table 1. Here, we mean by 'defining' that the respective connective is determined uniquely up to provable equivalence; that is, we have

Definition 3.1. Two Σ-sentences ϕ, ψ are *provably equivalent* if $\phi \vdash_\Sigma \psi$ and $\psi \vdash_\Sigma \phi$.

Proposition 3.2. The connectives $\vee$, $\wedge$, $\rightarrow$, $\top$, $\bot$ are determined uniquely up to provable equivalence by the respective properties shown in Table 1.

Proof We do only the case for $\vee$, the other cases being very similar. Let a Σ-sentence ρ satisfy the defining property of $\phi \vee \psi$ for Σ-sentences ϕ, ψ. Then $\rho \vdash_\Sigma \phi \vee \psi$ because $\phi \vdash_\Sigma \phi \vee \psi$ and $\psi \vdash_\Sigma \phi \vee \psi$, which in turn follows from the defining property of $\phi \vee \psi$ because $\phi \vee \psi \vdash_\Sigma \phi \vee \psi$. The converse entailment is analogous. □

They properties in Table1 correspond directly to standard proof rules; the 'only if' direction for implication $\rightarrow$ is known moreover as the *deduction theorem*. Importantly, the defining property of falsum $\bot$ is commonly referred to as *ex falso quodlibet*. 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.

Table 1 Properties of proof-theoretic connectives

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 $\Gamma \vdash \neg\varphi$

Next, we introduce *semantic* connectives. We say that a logic *has a semantic connective* if the corresponding sentences exist for defining properties analogous to the ones given in Table 1 but with entailment $\vdash$ replaced with logical consequence $\models$. It is then clear that the analogue of Proposition 3.2 holds, i.e. that semantic connectives are essentially unique when they exist, where 'essentially' means up to logical equivalence in the obvious sense:

Definition 3.3. Two Σ-sentences ϕ, ψ are *logically equivalent* if $\phi \models_\Sigma \psi$ and $\psi \models_\Sigma \phi$.

The following is clear:

Proposition 3.4. If $\mathcal{L}$ is sound and complete, then proof-theoretic and semantic connectives coincide (i.e. one exists iff the other does, and in this case they are the same).

Instead of defining semantic connectives via logical consequence, we may consider a stronger definition inspired by Tarski semantics, i.e. phrased in terms of satisfaction in a given model. The corresponding defining conditions are given in Table 2. A logic is said to *have a strong semantic connective* if it is possible to define an operation on sentences with the specified properties. It is easy to see the following.

Proposition 3.5. Every strong semantic connective is a semantic connective.

The converse does not hold. E.g. both **CPL** and **IPL-HA** have all semantic connectives; due to completeness, these coincide with the proof-theoretic ones. However, while all these connectives are strong in **CPL**, in **IPL-HA** only conjunction, truth, and falsum are strong. It is an interesting question whether there is a natural example of a logic that has semantic connectives that are neither strong nor, via completeness, inherited from the proof theory.

Table 2 Properties of strong semantic connectives

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 $\bot$	$M \not\models \bot$

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 4.1. A theory is unsatisfiable if it has no models.

Clearly, if $\mathcal{L}$ has a semantic falsum $\bot$ then T is unsatisfiable iff $T \models \bot$. However, there is more to inconsistency than just unsatisfiability[8], as witnessed by the historical development of research on inconsistency. According to Aristotle, inconsistency means that both some sentence as well as its negation can be proved:

Definition 4.2. Assume that $\mathcal{L}$ has a distinguished connective $\neg$ (not necessarily being a proof-theoretic or semantic negation[9]), and let T be a theory in $\mathcal{L}$. We say that T is *Aristotle inconsistent* if there is some sentence φ such that $T \vdash \varphi$ and $T \vdash \neg\varphi$.

This notion has several disadvantages. Firstly, it presupposes a notion of negation, which is not available in all logics. More importantly, it classifies paraconsistent logics as inconsistent, although of course paraconsistent logics were not known at Aristotle's time.

A modern definition of inconsistency overcoming this problem was coined by David Hilbert. Hilbert was the initiator of the famous *Hilbert program*, 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 4.3 (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 was doomed to be unsuccessful. As a result, the question whether theories like ZFC (which is used as the 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. But all we have so far in this respect are relative results, such as another famous result by Gödel:

ZFC is inconsistent iff ZF is inconsistent[11],

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.

---

[8] Even though the Rolling Stones prioritize unsatisfiability.

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

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

[11] Actually, Gödel proved the corresponding statement about unsatisfiability, but by Gödel's completeness theorem for first-order logic, inconsistency and unsatisfiability are equivalent here; see also our Prop. 5.3 below.

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

Definition 4.4 (Hilbert). Assume that $\mathcal{L}$ has a distinguished constant $\bot$ (not necessarily being a proof-theoretic or semantic falsum[12]). Then T is $\bot$-*inconsistent* if $T \vdash \bot$.

Still, this definition does not work with logics that do not have $\bot$, for example positive logic or equational logic. Hilbert therefore also proposed a notion of inconsistency that has no prerequisites (i.e. no logical connectives are needed) and simultaneously is most powerful one among the notions considered so far, in terms of the logical strength of inconsistent theories:

Definition 4.5 (Hilbert). A theory T over Σ is *absolutely inconsistent* if $T \vdash \varphi$ for all Σ-sentences φ.

This definition abstracts the principle of *ex falso quodlibet*, without however postulating the existence of $\bot$.

Example 4.6. In equational logic, the theory $\{f(x,y) = y, f(x,y) = f(y,x)\}$ is absolutely inconsistent (we have $x = f(y,x) = f(x,y) = y$, so that every equation is derivable). None of the other notions of inconsistency mentioned so far apply here, as equational logic has neither negation nor falsum.

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 obtain such a variable). Unfortunately, this notion is too closely tied to a specific logical system to be of interest here. [13]

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

Proposition 4.7.

1. If $\mathcal{L}$ has a constant $\bot$ then absolute inconsistency implies $\bot$-inconsistency, and if $\mathcal{L}$ has an operation $\neg$ then absolute inconsistency implies Aristotle inconsistency.
2. In the presence of proof-theoretic falsum $\bot$, absolute inconsistency and $\bot$-inconsistency are equivalent.
3. In the presence of proof-theoretic falsum and negation, all three notions of inconsistency are 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 absolute inconsistency. By the definition of proof-theoretic negation, from $\Gamma \vdash \neg\varphi$ we obtain $\Gamma \cup \{\varphi\} \vdash \bot$. Together with $\Gamma \vdash \varphi$, this leads to $\Gamma \vdash \bot$. □

[12] 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.

[13] Given mild assumptions on the signature category it would be rather easy to come up with a general formulation of this notion, e.g. saying that T derives a non-trivial (i.e. not everywhere satisfied) sentence in a subsignature that is disjoint from a subsignature supporting T. However, in order to be really useful, also some notion of substitution is needed, resulting in a rather complicated definition.

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

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 5.1. In presence of proof-theoretic negation, falsum and truth,

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

Here, we use $\dashv\vdash$ to denote mutual entailment.

We say that proof-theoretic negation is *classical*, if $\neg\neg\varphi \vdash \varphi$.

Proposition 5.2. Proof-theoretic negation is classical in complete logics with strong semantic negation.

Proof By strong semantic negation, $\neg\neg\varphi \models \varphi$, hence by completeness, $\neg\neg\varphi \vdash \varphi$. □

Soundness and completeness, while defined in terms of entailment, can be characterized completely in terms of inconsistency and unsatisfiability. Recall from Proposition 4.7 that all notions of inconsistency are equivalent in logics with proof-theoretic falsum and negation, so we just use the unspecific term *inconsistent* in the following statement.

Proposition 5.3. Let $\mathcal{L}$ be a logic with both proof-theoretic and semantic negation, truth and falsum, such that proof-theoretic negation is classical. Then

1. $\mathcal{L}$ is sound if and only if every inconsistent theory in $\mathcal{L}$ is unsatisfiable.
2. $\mathcal{L}$ is complete if and only if every unsatisfiable theory in $\mathcal{L}$ is inconsistent.

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

(1), "$\Leftarrow$" 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), "$\Rightarrow$" Let T be unsatisfiable, i.e. $T \models \bot$. By completeness, $T \vdash \bot$, hence T is inconsistent.

(2), "$\Leftarrow$" Let $T \models \varphi$. Then $T \cup \{\neg\varphi\}$ is not satisfiable, and hence inconsistent by the assumption. From $T \cup \{\neg\varphi\} \vdash \bot$, we obtain $T \vdash \neg\neg\varphi$ and hence by classicality $T \vdash \varphi$. □

It should be stressed that these proofs, which avoid any form of negated relations such as $\nvdash$ or $\not\models$, become less elegant when one reformulates them in terms of consistency and satisfiability, as some over-cautious logicians do — logicians tend to be easily frightened by inconsistencies[15]. The more natural relation is indeed that between inconsistency and unsatisfiability.

Definition 5.4. A logic is *paraconsistent* if it has a negation operator for which Aristotle inconsistency does not imply absolute inconsistency. Such an operator is then called a *paraconsistent negation*.

Example 5.5. Belnap's four-valued logic[1] has, in the base version, the same syntax as propositional logic. It evaluates formulas over four truth values t (*true*), f

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

(*false*), $\ddot{\bot}$ (*unknown*), and $\ddot{\top}$ (*contradictory*); i.e. models are maps from the set of propositional variables into $4 = \{t, f, \ddot{\bot}, \ddot{\top}\}$, and sentences are evaluated recursively to a truth value in 4 using prescribed truth tables for the connectives. Negation $\neg$ works as usual on t and f, and has $\ddot{\bot}$ and $\ddot{\top}$ as fixed points. Of course, the constant falsum (still written $\bot$ in the logical syntax) for use in $\bot$-inconsistency evaluates to f, not $\ddot{\bot}$. Both t and $\ddot{\top}$ are designated truth values, i.e. for a Σ-model M and a Σ-sentence φ, $M \models_\Sigma \varphi$ iff φ evaluates to either t or $\ddot{\top}$ in M. Thus, the theory $T = \{a, \neg a\}$ (for some propositional variable a) is Aristotle inconsistent but not, assuming any sound entailment system, $\bot$-inconsistent (and hence not absolutely inconsistent), as $T \not\models_\Sigma \bot$: By assigning the truth value $\ddot{\top}$ to a, we obtain a model M with $M \models_\Sigma T$ but not $M \models_\Sigma \bot$. Thus, $\neg$ is indeed a paraconsistent negation.

Proposition 5.6.

1. A paraconsistent negation cannot be proof-theoretic.
2. In a sound and complete logic, paraconsistent negation cannot be semantic.

Proof 1. By Prop. 4.7.

2. In a sound and complete logic, a semantic negation is also a proof-theoretic negation. □

6 Conservative Extensions

In the structured development of theories, the notion of conservativity plays a crucial role[2]. Like many other concepts considered here, conservativity comes in two flavours, a syntactic and a semantic one:

Definition 6.1. A theory morphism $\sigma\colon T_1 \longrightarrow T_2$ is

1. *model-theoretically conservative* if each T_1-model M_1 has a *σ-expansion* to a T_2-model M_2, i.e. $M_2|_\sigma = M_1$;
2. *consequence-theoretically conservative* if for each sentence φ of the same signature as T_1,
$$T_2 \models \sigma(\varphi) \text{ implies } T_1 \models \varphi;$$
3. *proof-theoretically conservative* if the same holds for $\vdash$, i.e. for each sentence φ of the same signature as T_1,
$$T_2 \vdash \sigma(\varphi) \text{ implies } T_1 \vdash \varphi.$$

The relation between these notions is as follows:

Proposition 6.2. 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 reducing inconsistency (resp. unsatisfiability) of a smaller theory to that of larger one, using the following obvious result:

Proposition 6.3.

1. Consequence-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 potentially easier to prove that they are inconsistent. (Indeed, the largest theory, consisting of *all* sentences, is always absolutely 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 summarized and discussed recent research aimed at proving inconsistency of specifications in a structured and institution-independent way. These efforts form a 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[15])
- the existence of natural numbers $(a, b, x, y) \neq (2, 3, 3, 1)$ such that $x^a - y^b = 1$ (Preda Mihăilescu[10])
- the existence of a simply connected closed 3-manifold not homeomorphic to S^3 (Grigori Perelman[11,12,13])

Moreover, recently completed work in automated theorem proving, Thomas Hales' FlysPecK project (http://code.google.com/p/flyspeck/), has been directed at reducing the inconsistency of ZF to the 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.

Acknowledgements

We thank Erwin R. Catesbeiana for cooperation on an earlier draft of this paper.

References

[1] Belnap N. A useful four-valued logic. Modern Uses of Multiple-Valued Logic. Reidel. 1977. 7–73.

[2] Codescu M, Mossakowski T, Maeder C. Checking conservativity with Hets. In: Heckel R, Milius S, eds. CALCO 2013. Lecture Notes in Computer Science. Springer-Verlag Berlin, Heidelberg. 2013, 8089: 315–321.

[3] Diaconescu R. Institution-Independent Model Theory. Birkhäuser, 2008.

[4] Fiadeiro J, Sernadas A. Structuring theories on consequence. In: Sannella D, Tarlecki A, eds. 5th WADT, LNCS. Springer Verlag, 1988, 332: 44–72.

[5] Goguen JA, Burstall RM. Institutions: Abstract model theory for specification and programming. Journal of the Association for Computing Machinery, 1992, 39: 95–146. LNCS, 1984, 164: 221–256.

[6] Herrlich H, Strecker G. Category Theory. Allyn and Bacon, Boston, 1973.

[7] Horrocks I, Voronkov A. Reasoning support for expressive ontology languages using a theorem prover. Proc. of the Fourth International Symposium on Foundations of Information and Knowledge Systems (FoIKS). Lecture Notes in Computer Science. Springer, 2006, 3861: 201–218.

[8] Lambek J, Scott PJ. Introduction to Higher Order Categorical Logic. Cambridge University Press, 1986.

[9] Meseguer J. General logics. Logic Colloquium. North Holland, 1989, 87: 275–329.

[10] Mihăilescu P. Primary cyclotomic units and a proof of Catalan's conjecture. J. Reine angew. Math., 2004, 572: 167–195.

[11] Perelman G. The entropy formula for the Ricci flow and its geometric applications. arXiv:math.DG/, 2002.

[12] Perelman G. Ricci flow with surgery on three-manifolds. arXiv:math.DG/0303109, 2002.

[13] Perelman G. Finite extinction time for the solutions to the ricci flow on certain three-manifolds. arXiv:math.DG/0211159, 2003.

[14] Troelstra A, Schwichtenberg H. Basic Proof Theory. Number 43 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1996.

[15] Wiles A. Modular elliptic curves and Fermat's last theorem. Ann. Math., 1995, 141: 443–551.